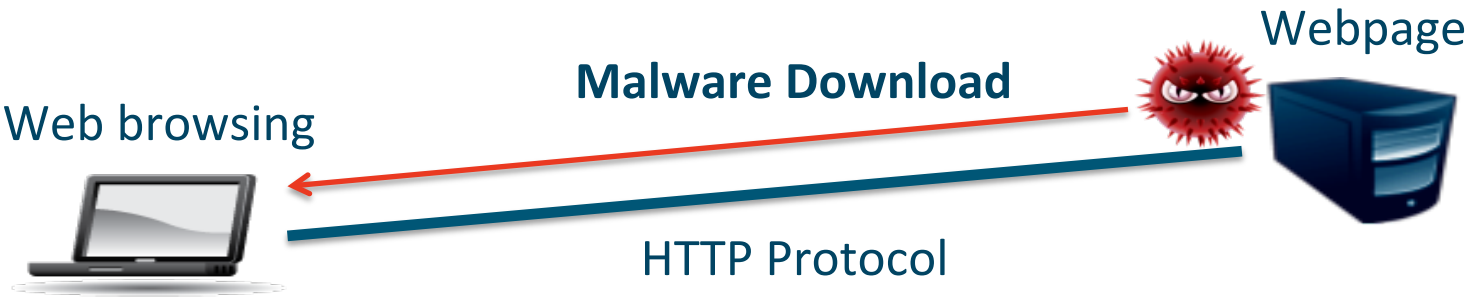
WebWitness: Investigating, Categorizing, and Mitigating Malware Download Paths

Terry Nelms[1,2], Roberto Perdici[3,2], Manos Antonakakis[2], Mustaque Ahamd[2,4]

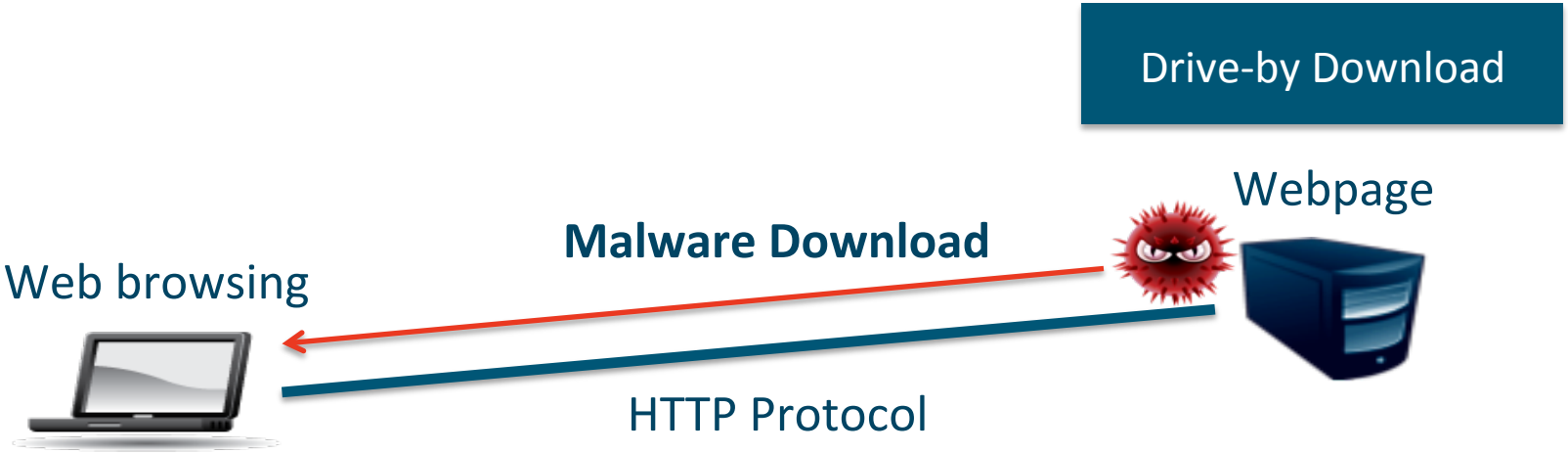
[1]Damballa, Inc.
[2]Georgia Institute of Technology
[3]University of Georgia
[4]New York University Abu Dhabi

tnelms@damballa.com, perdisci@cs.uga.edu, manos@gatech.edu, mustaq@cc.gatech.edu

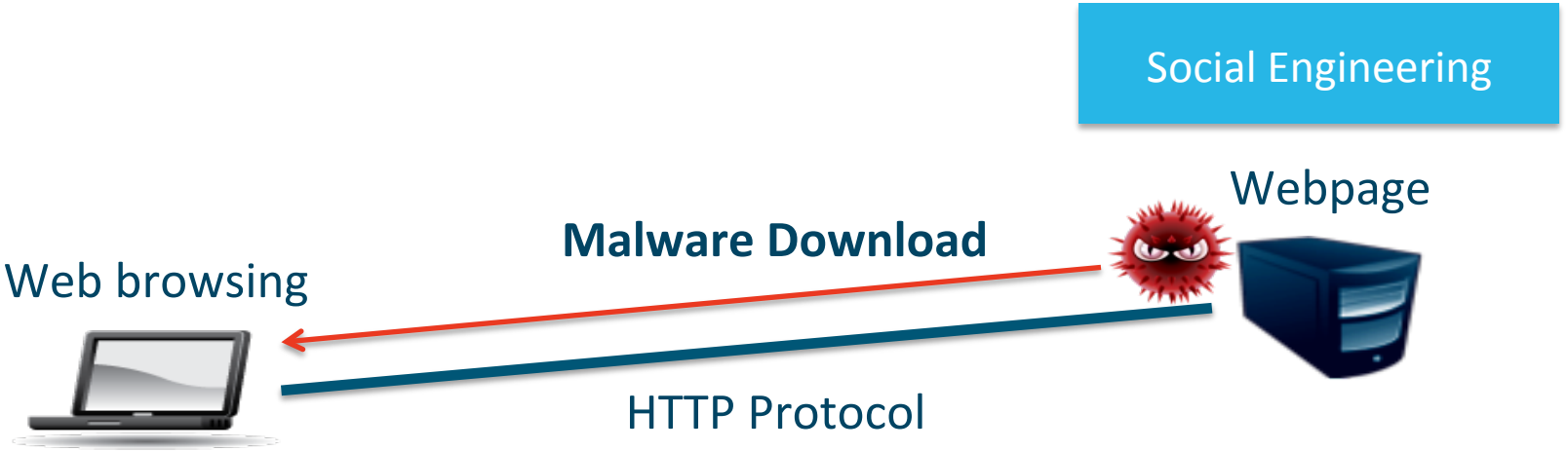
Web Browser – Popular Infection Vector



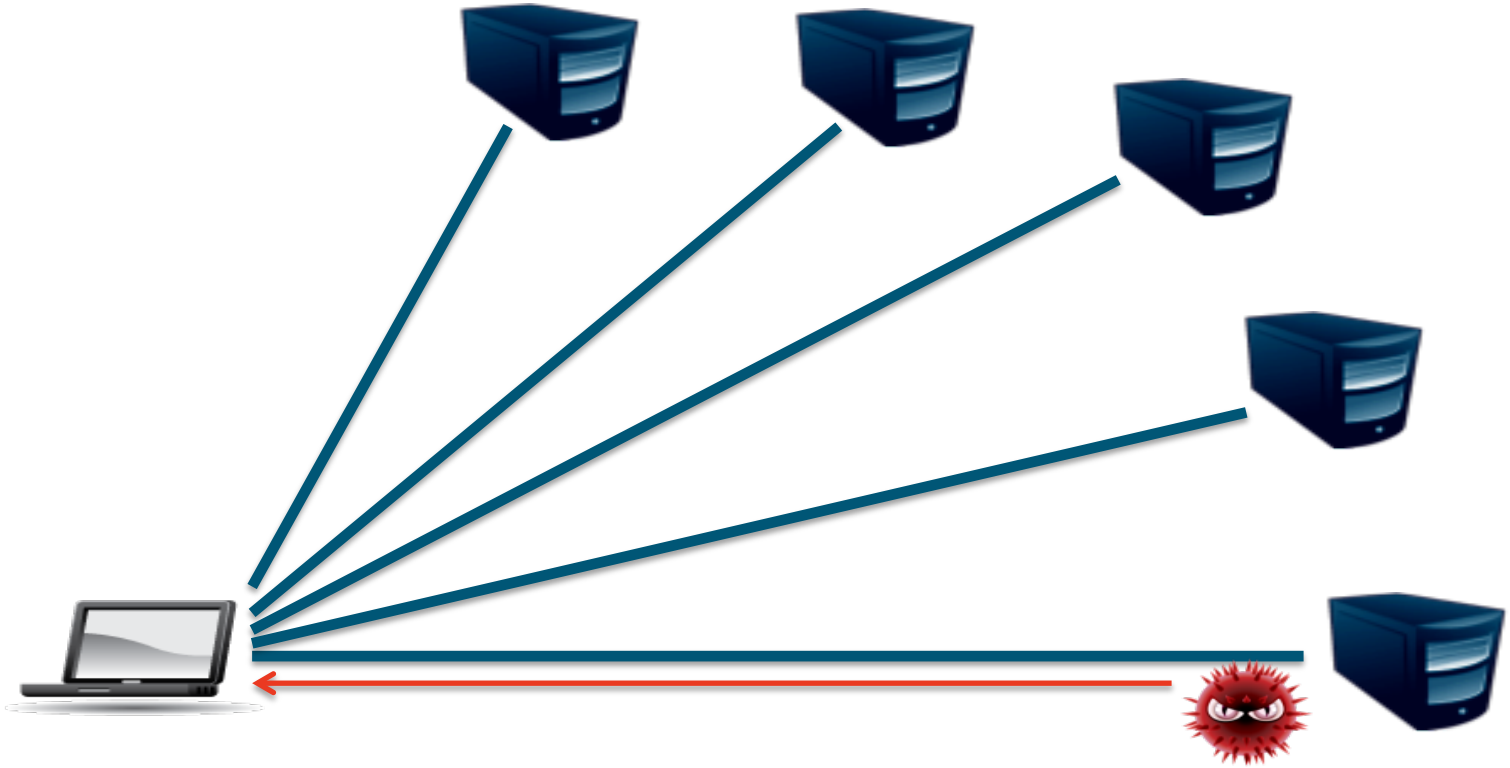
How do users get infected?



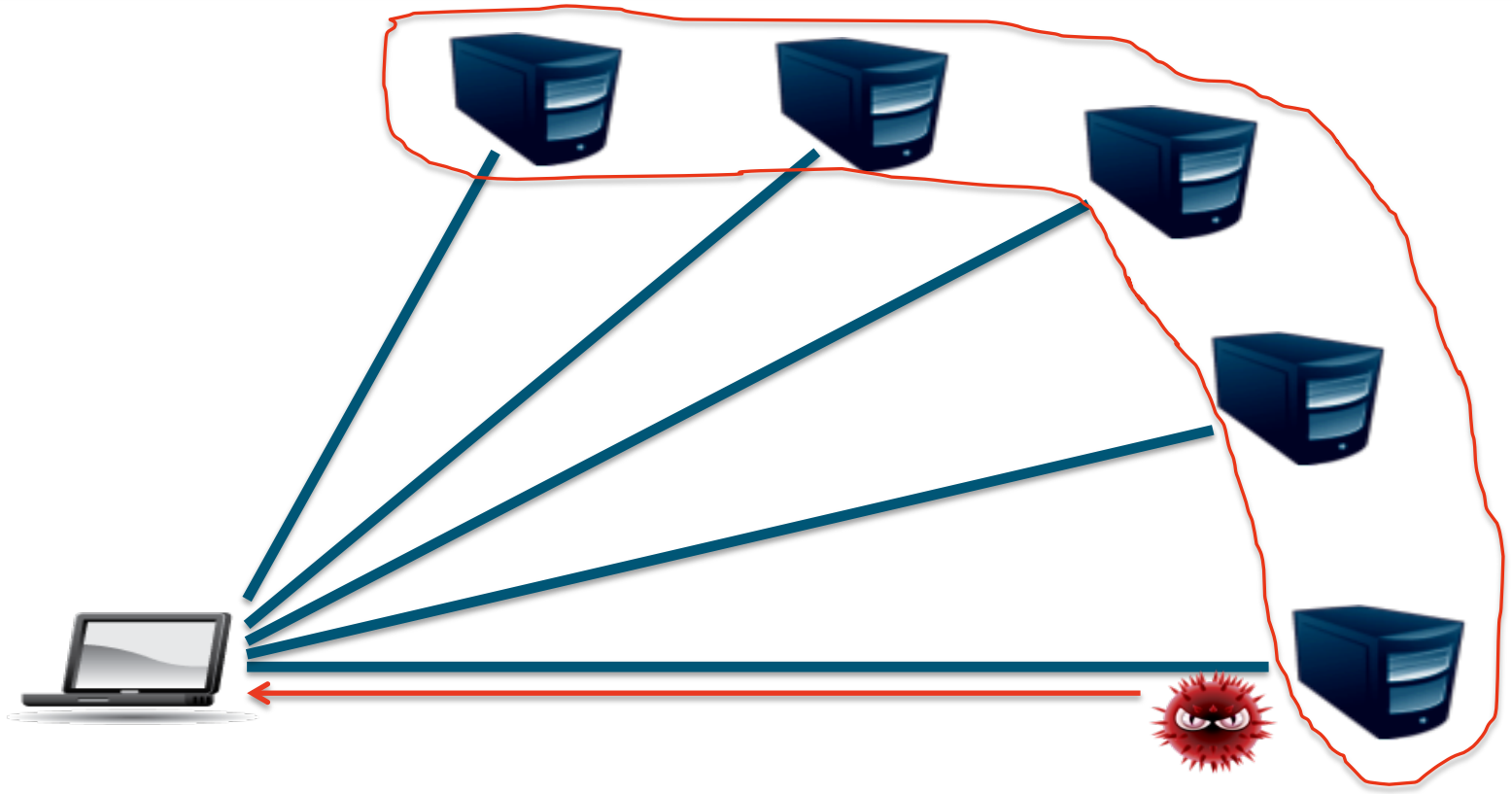
How do users get infected?



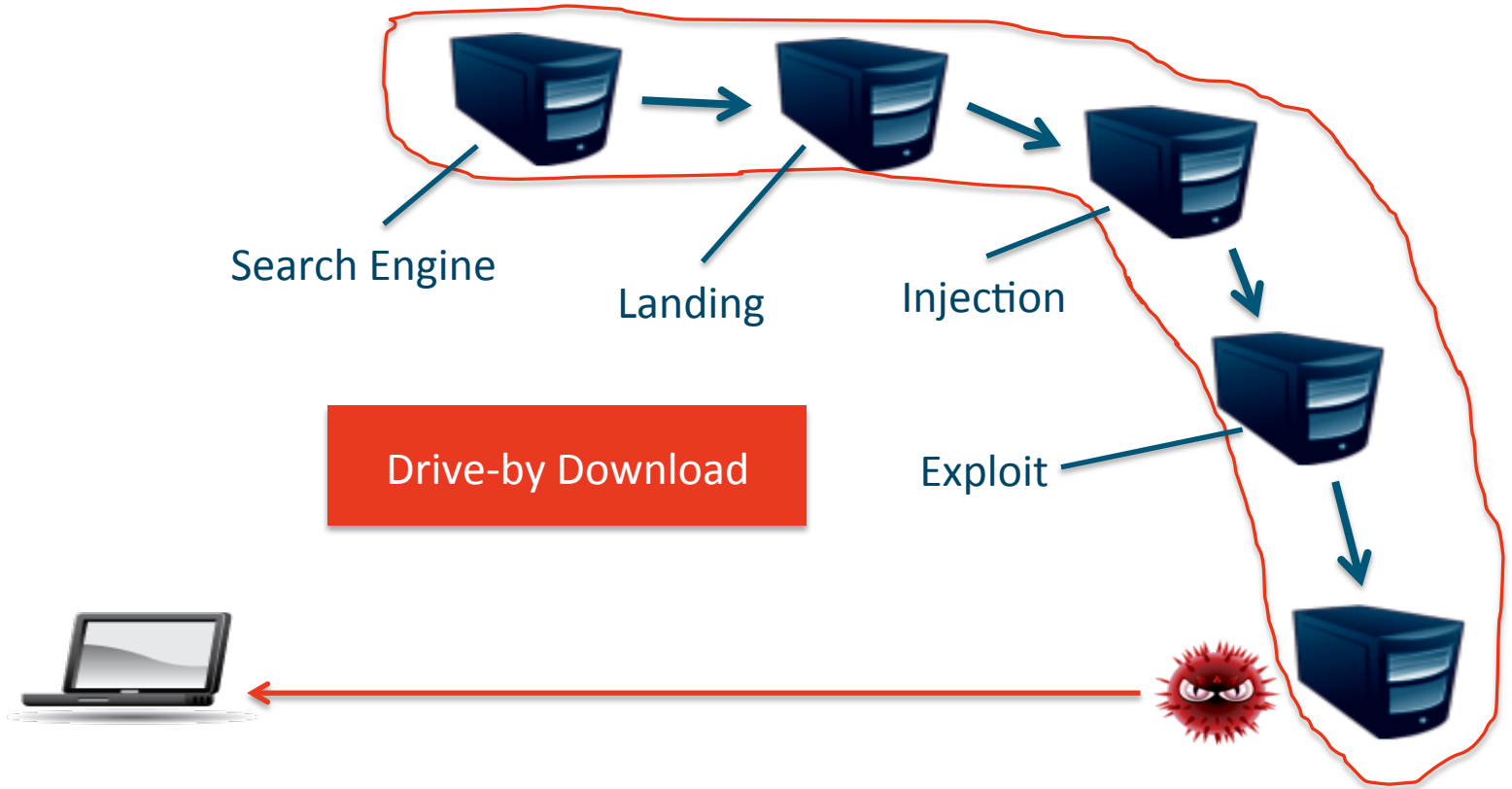
What happened before the download?



What happened before the download?



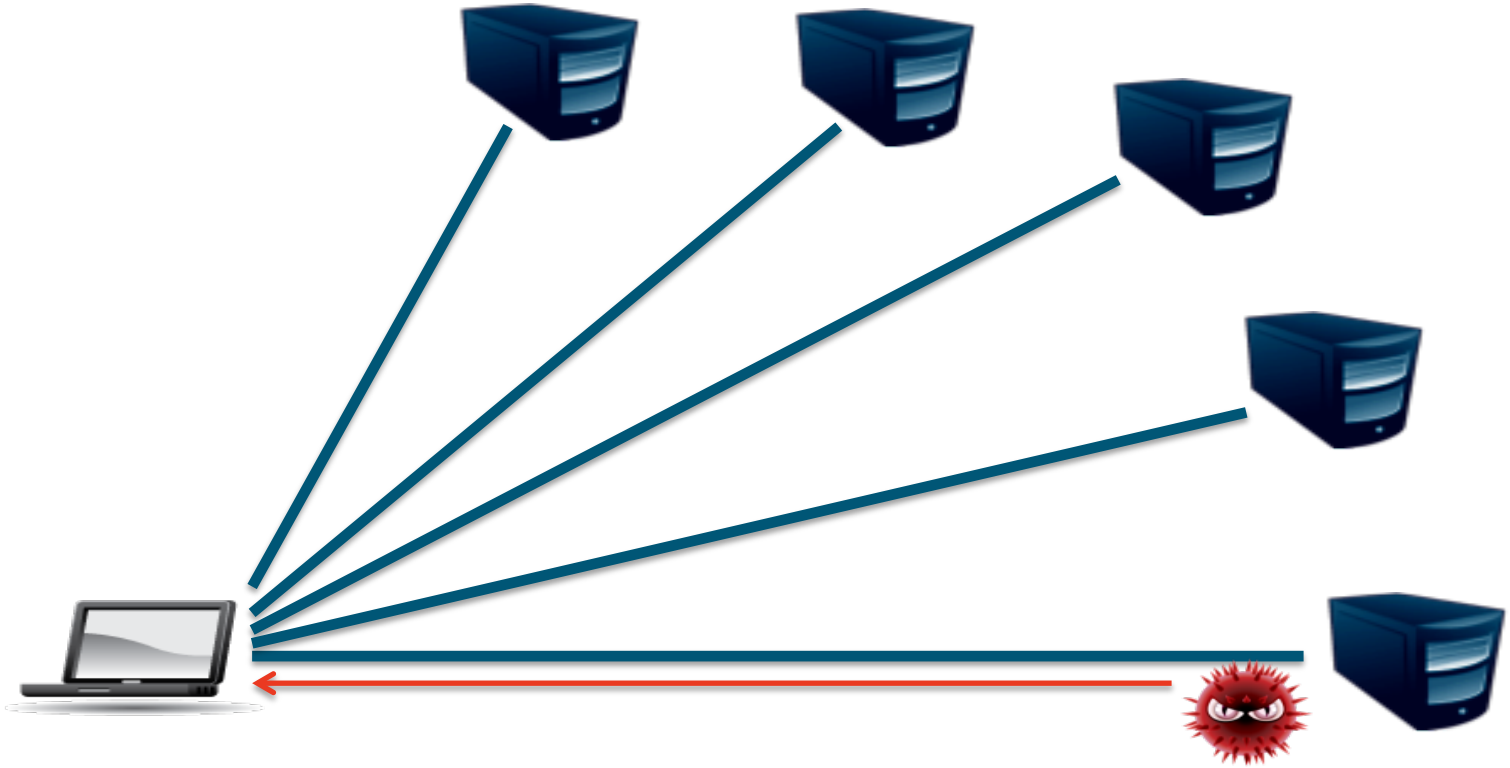
What happened before the download?



Our Work

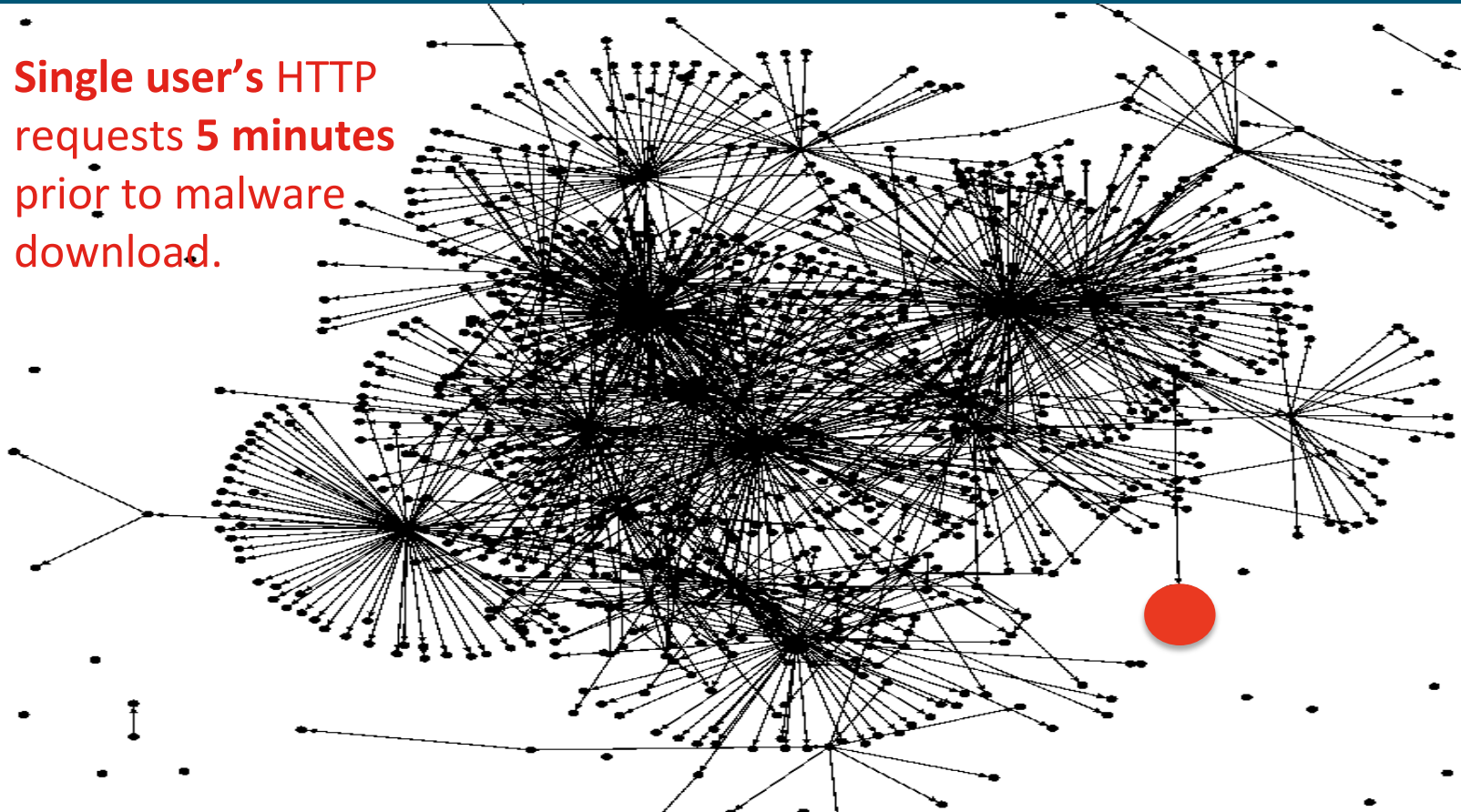
- Automatically provide context to attacks.
  - Reconstruct download path.
  - Identify download cause.
- Understand current attack trends.
- Develop more effective defenses.

How hard is it to find the download path?

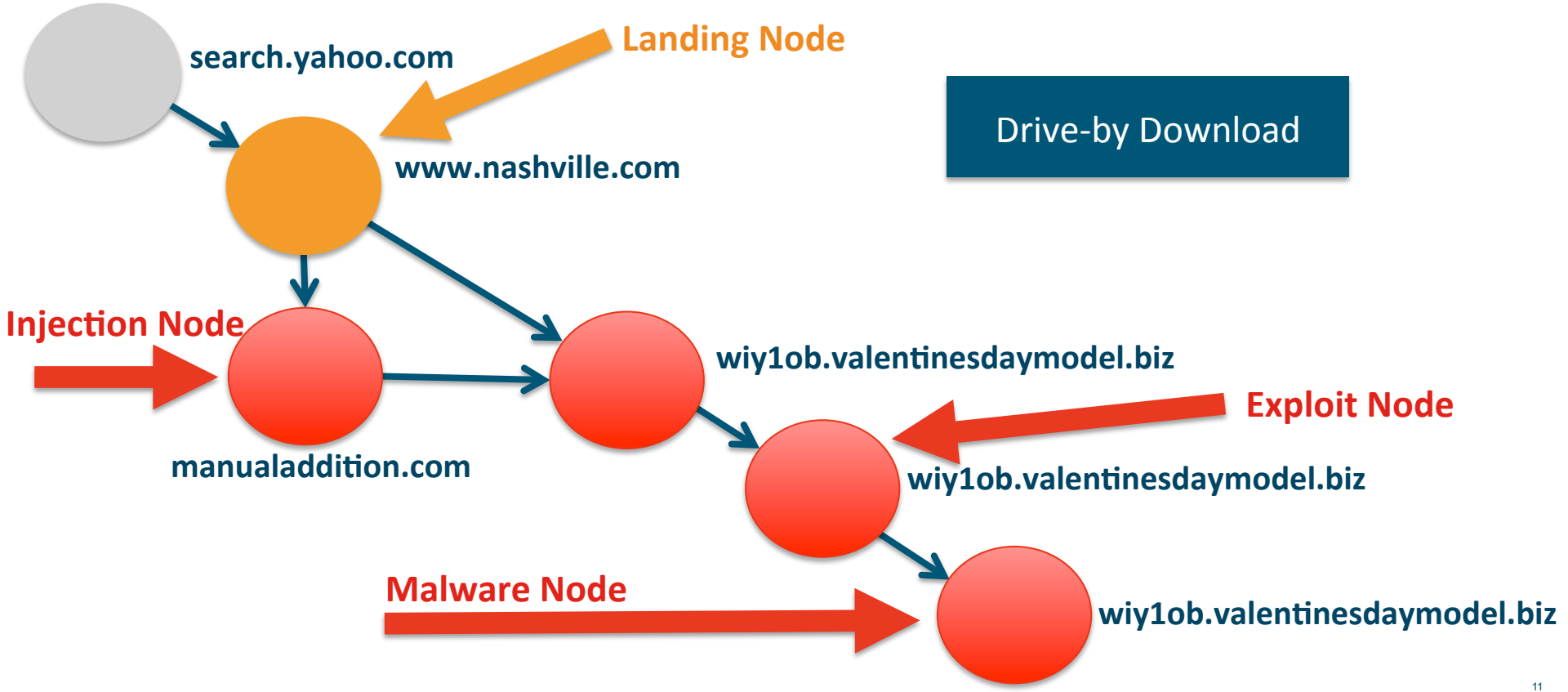


Web Traffic Graph

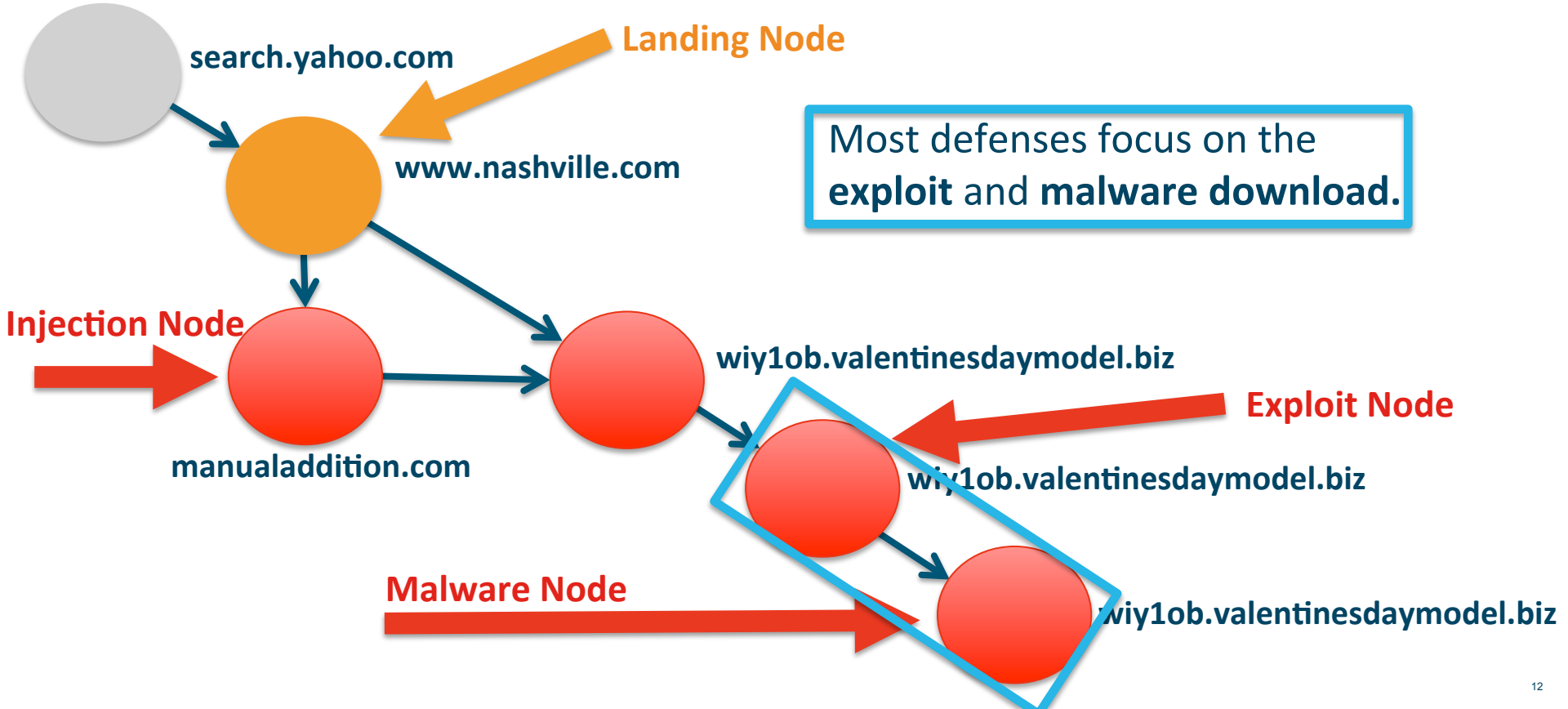
Single user's HTTP requests **5 minutes** prior to malware download.



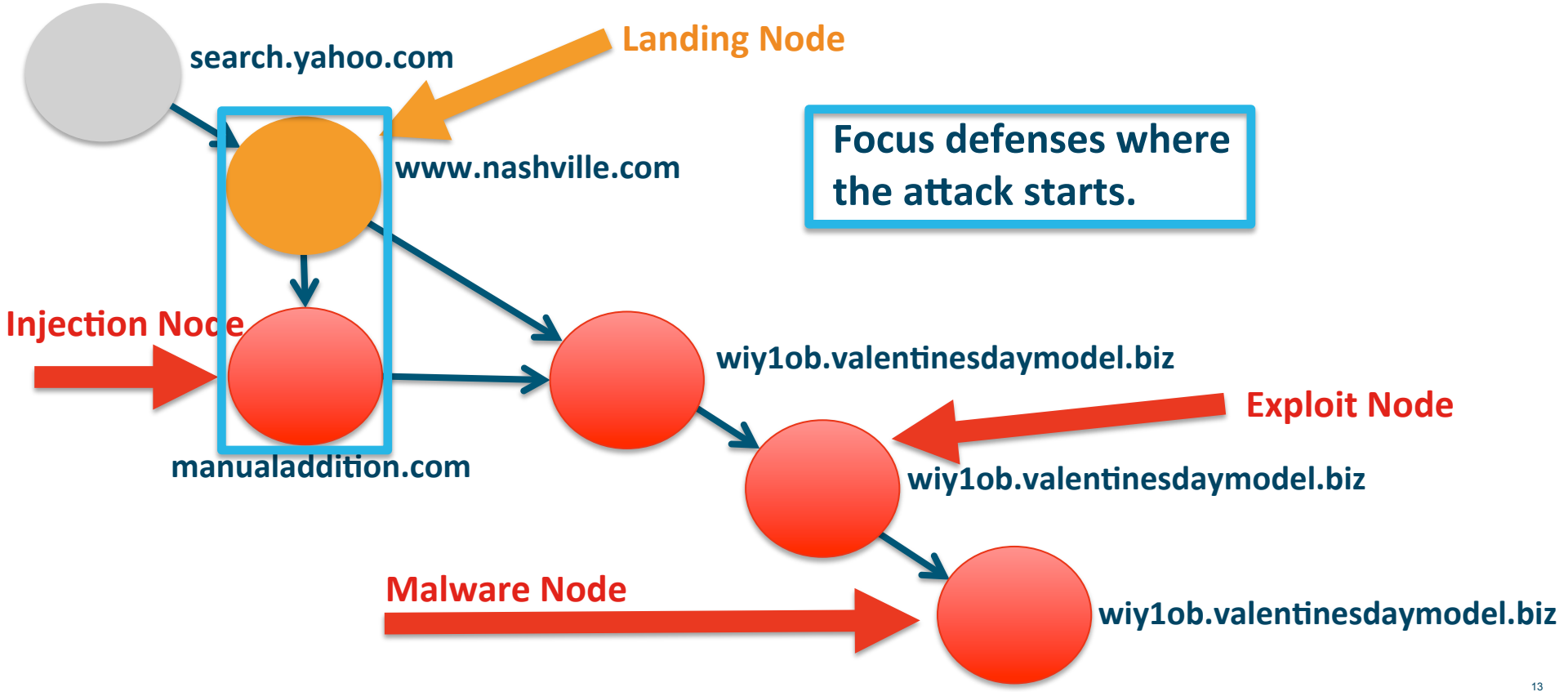
Annotated Drive-by Download Path



Drive-by Download Defense



Drive-by Download Defense



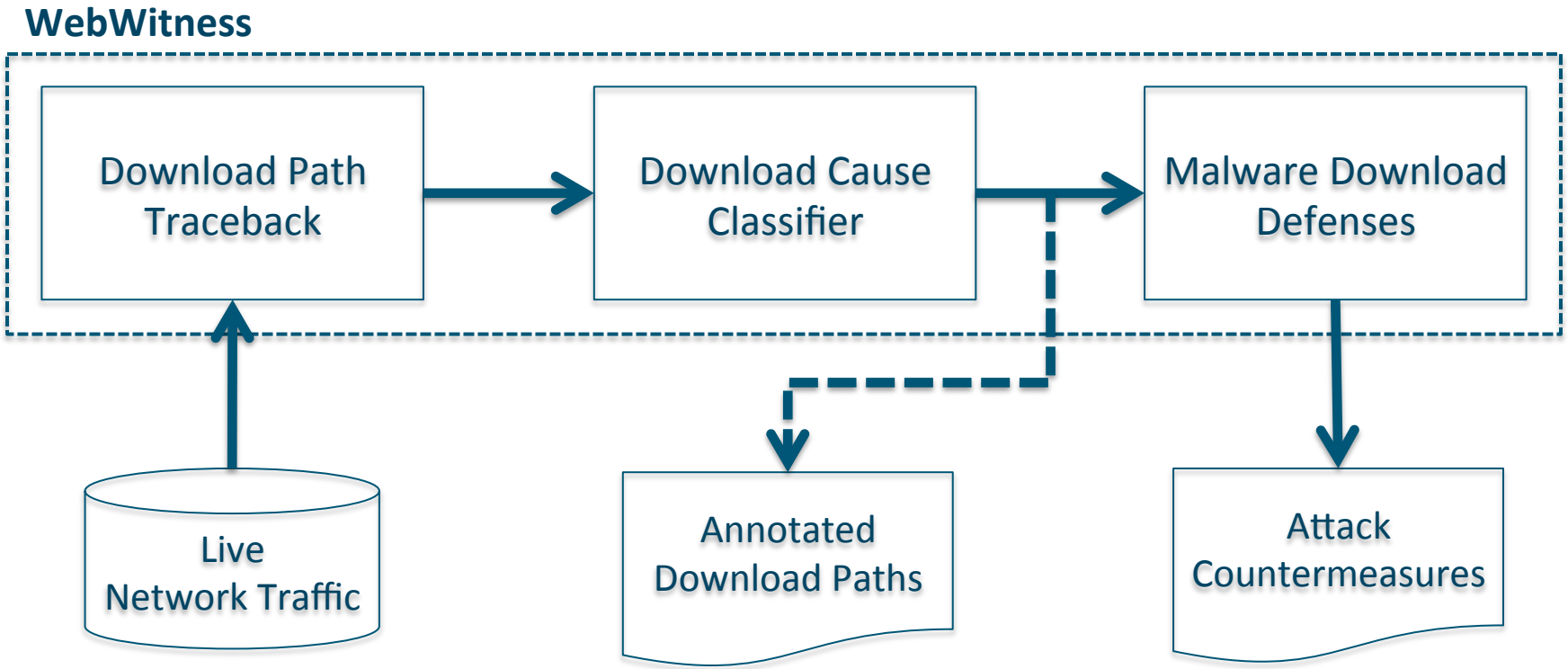
In-The-Wild Malware Study

- Collection agent deployed on large academic network for a period of six months.
- A total of 533 malware downloads due to drive-by, social engineering and update.
- Extensive manual analysis of malicious downloads.
  - Reverse engineer content.
  - Deobfuscate java script.

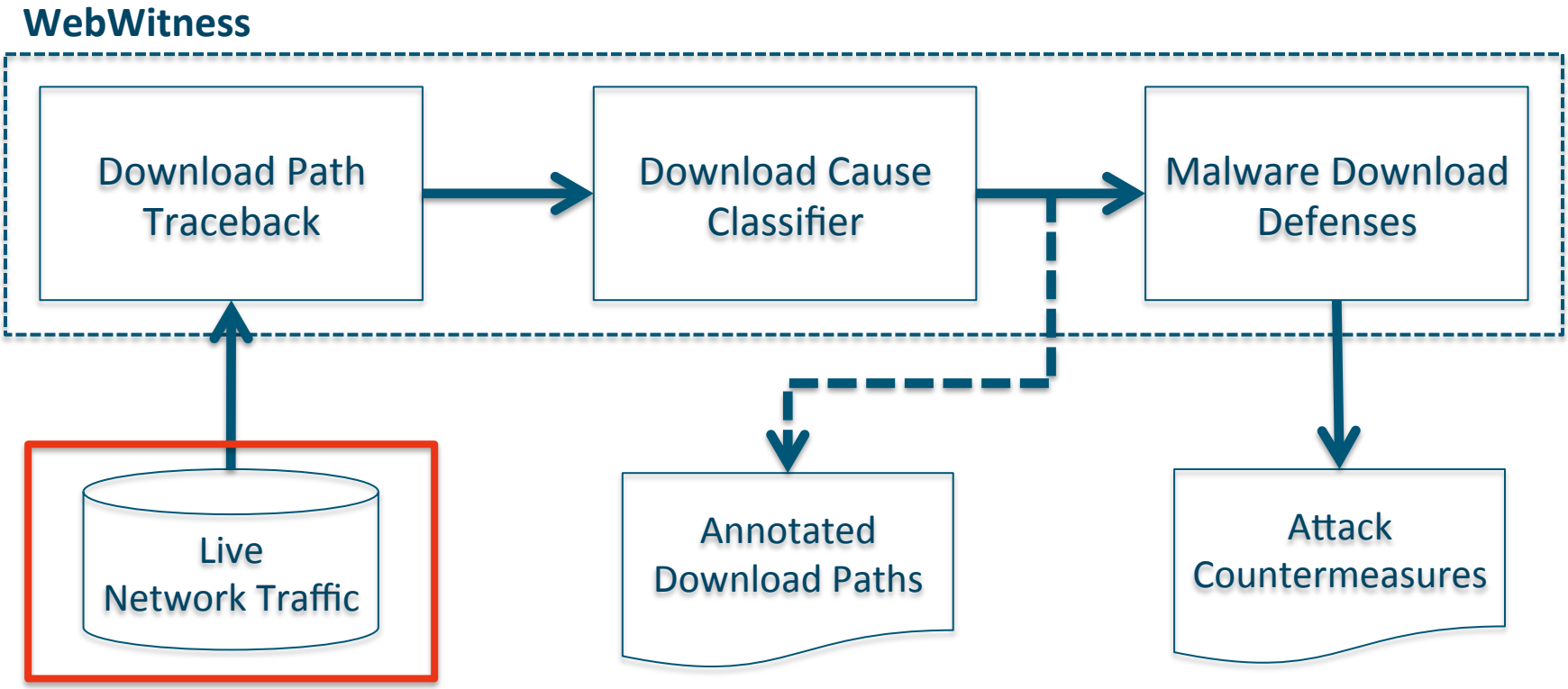
In-The-Wild Malware Study Results

- Download path trace back is difficult.
  - Cannot only rely on Referrer and Location headers.
- Labeling the role of nodes extremely challenging.
  - Semantic gap between network and browser.

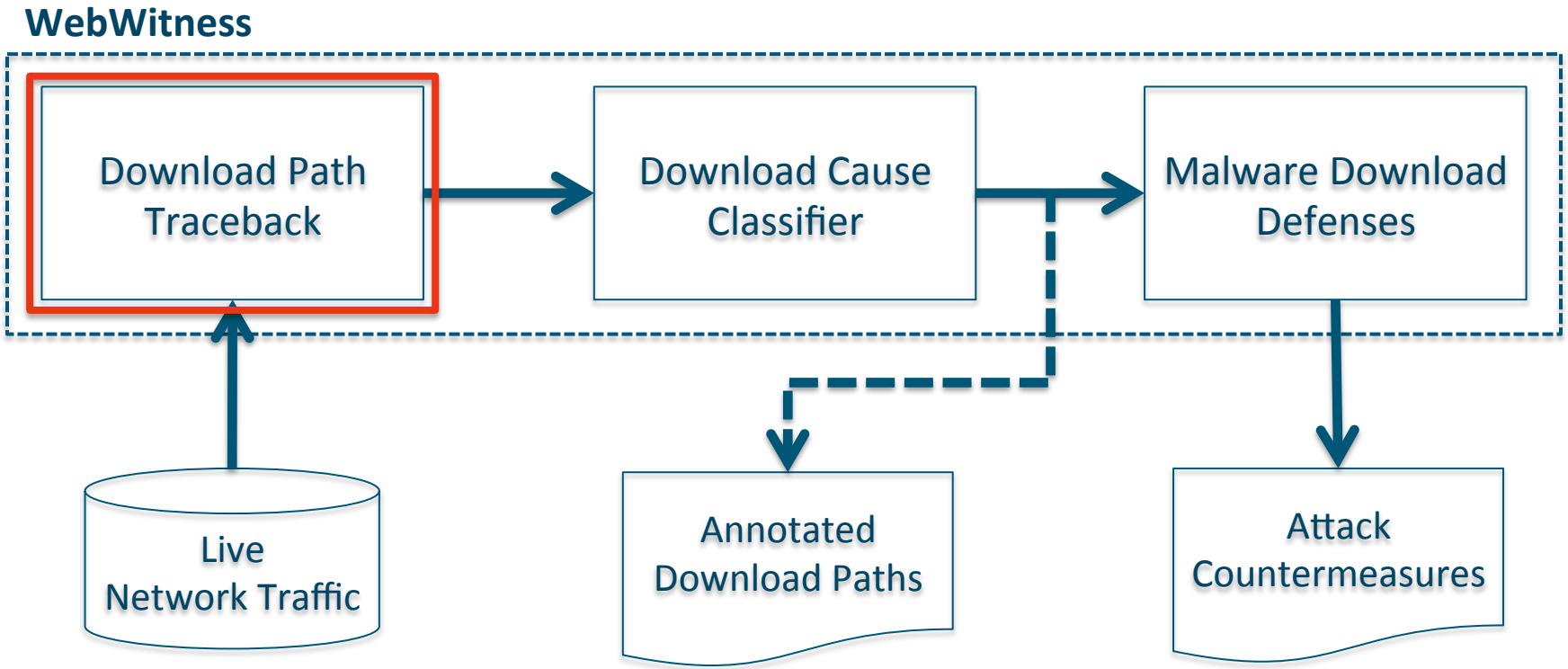
WebWitness System



WebWitness System

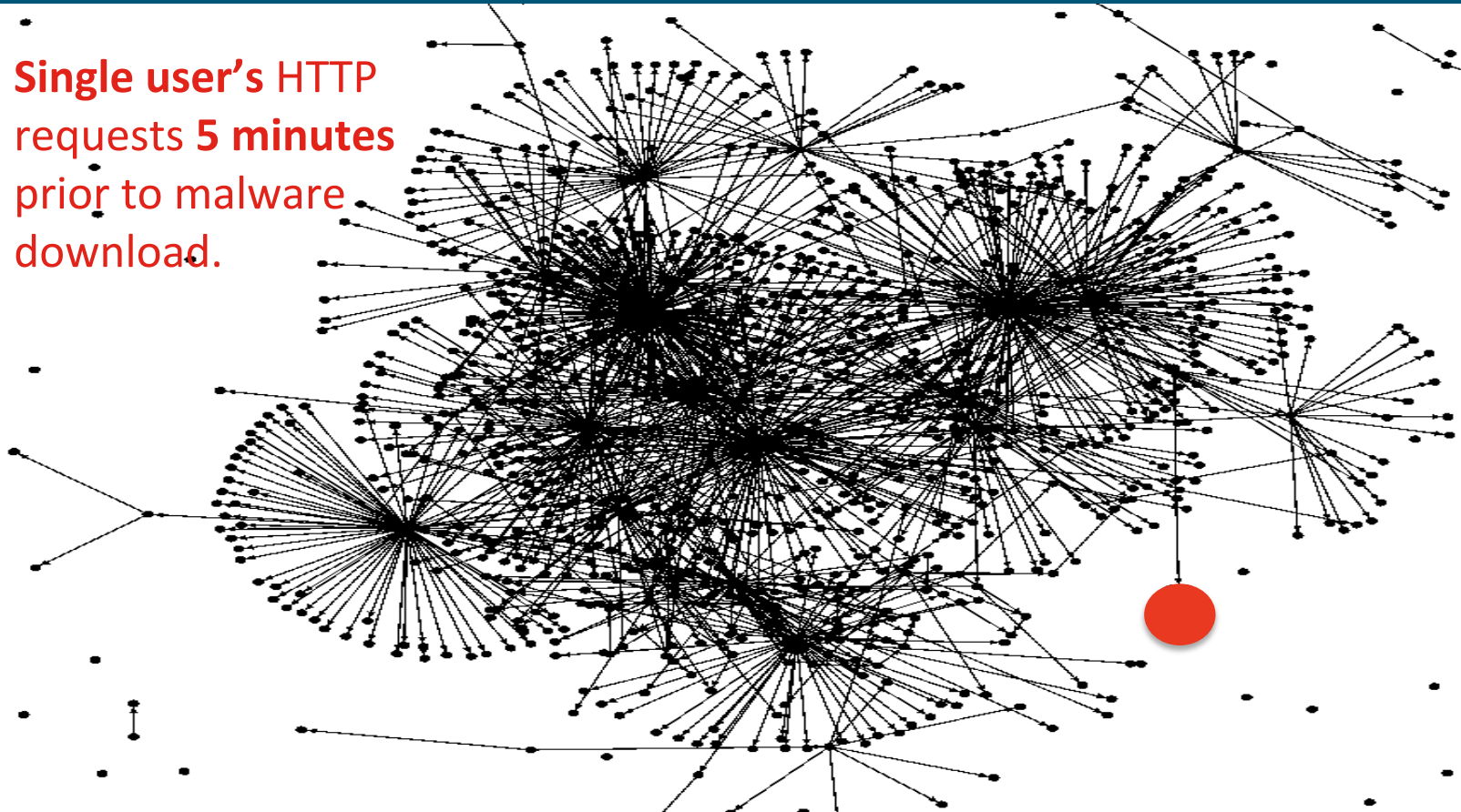


WebWitness System



Web Traffic Graph

Single user's HTTP requests **5 minutes** prior to malware download.



Referrer & Location Header Only

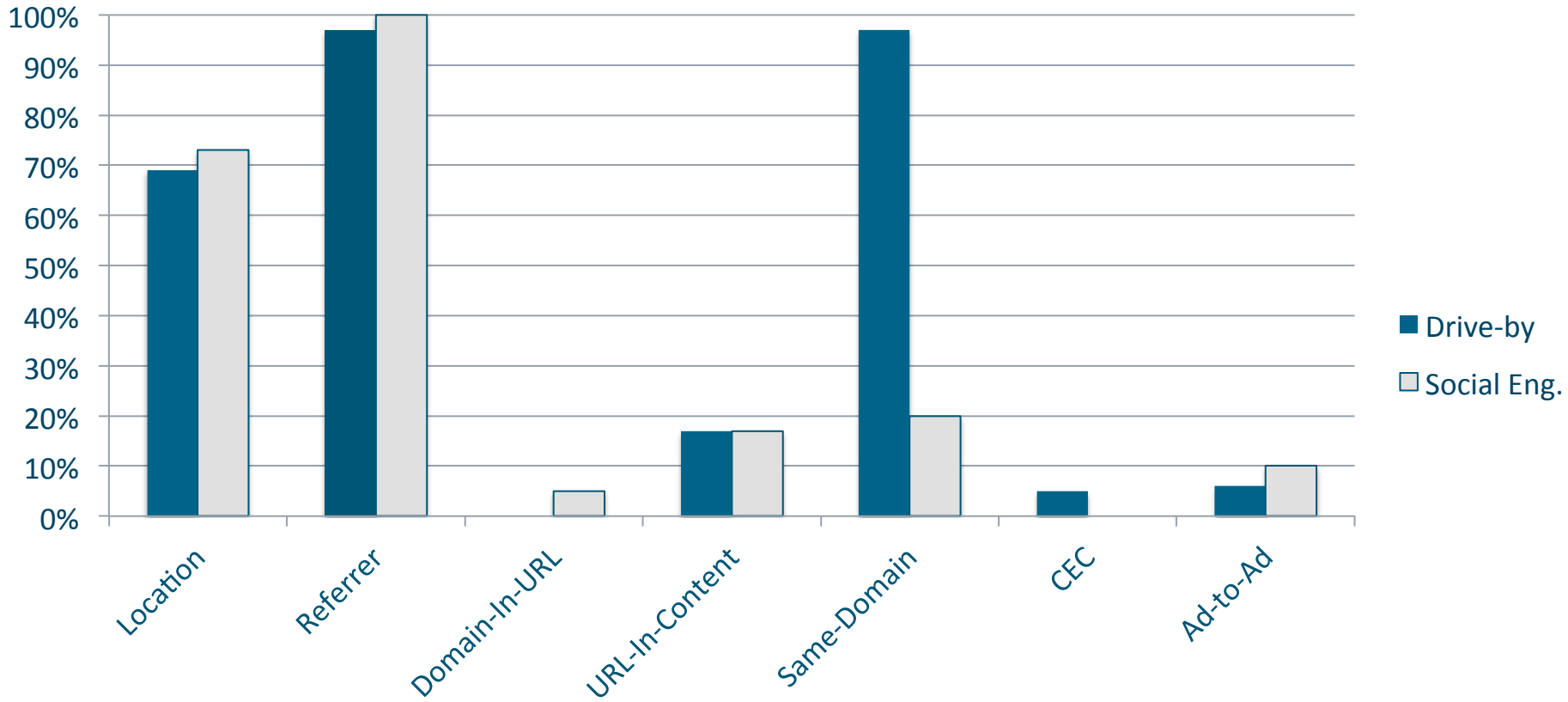
Drive-by

0%

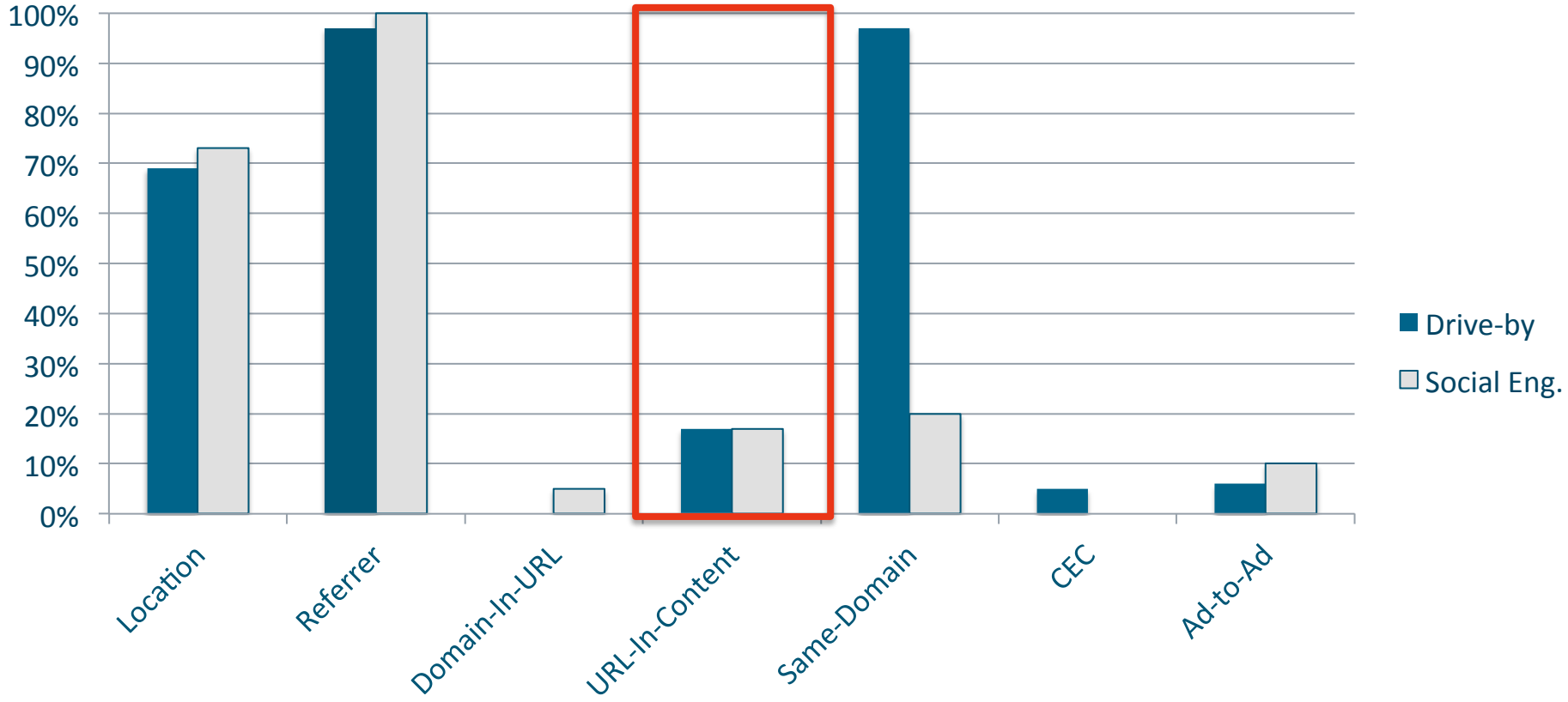
Social Eng.

53%

Web Graph “Source-of” Relationships



Web Graph “Source-of” Relationships



Improvement using all “source-of” features

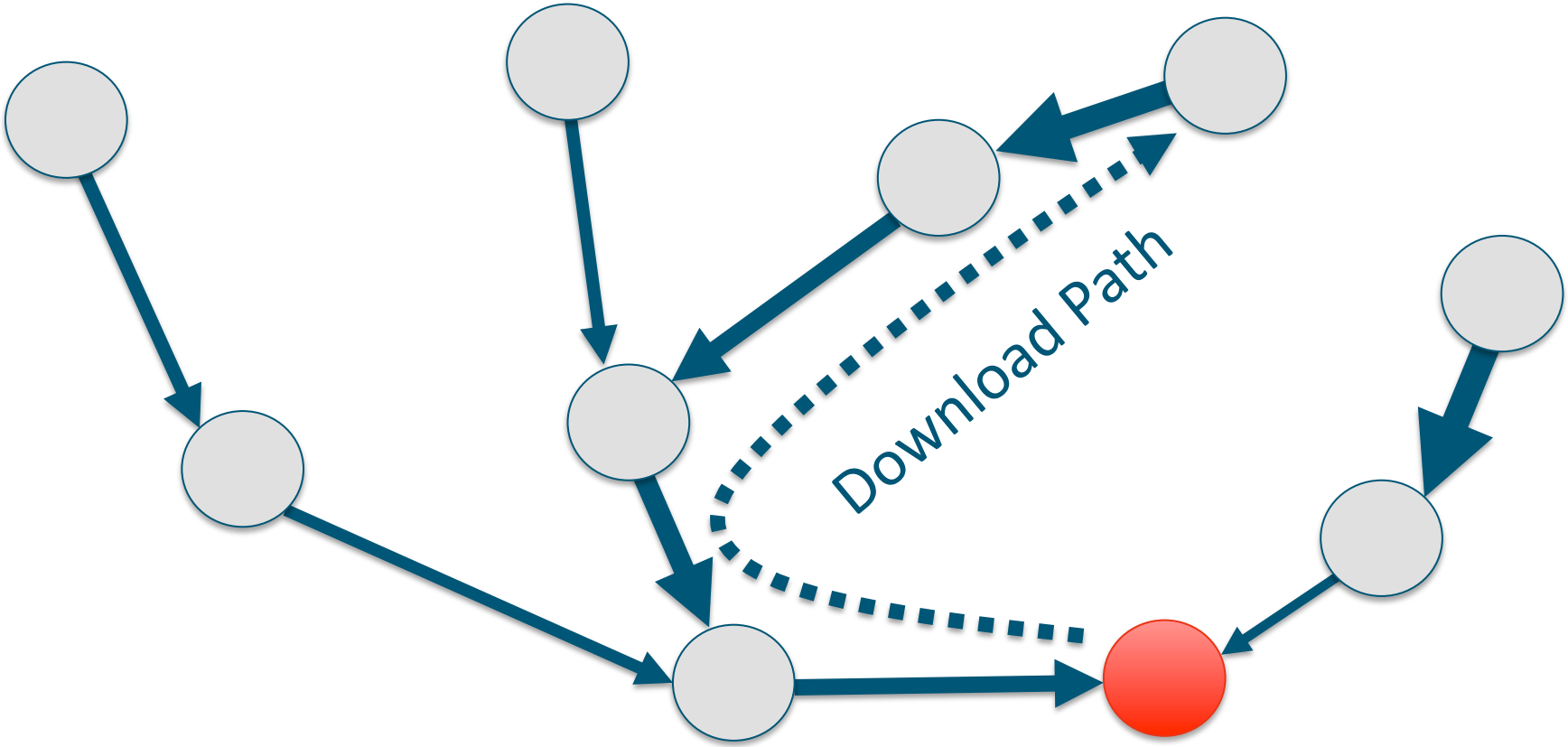
Drive-by

0% → **96%**

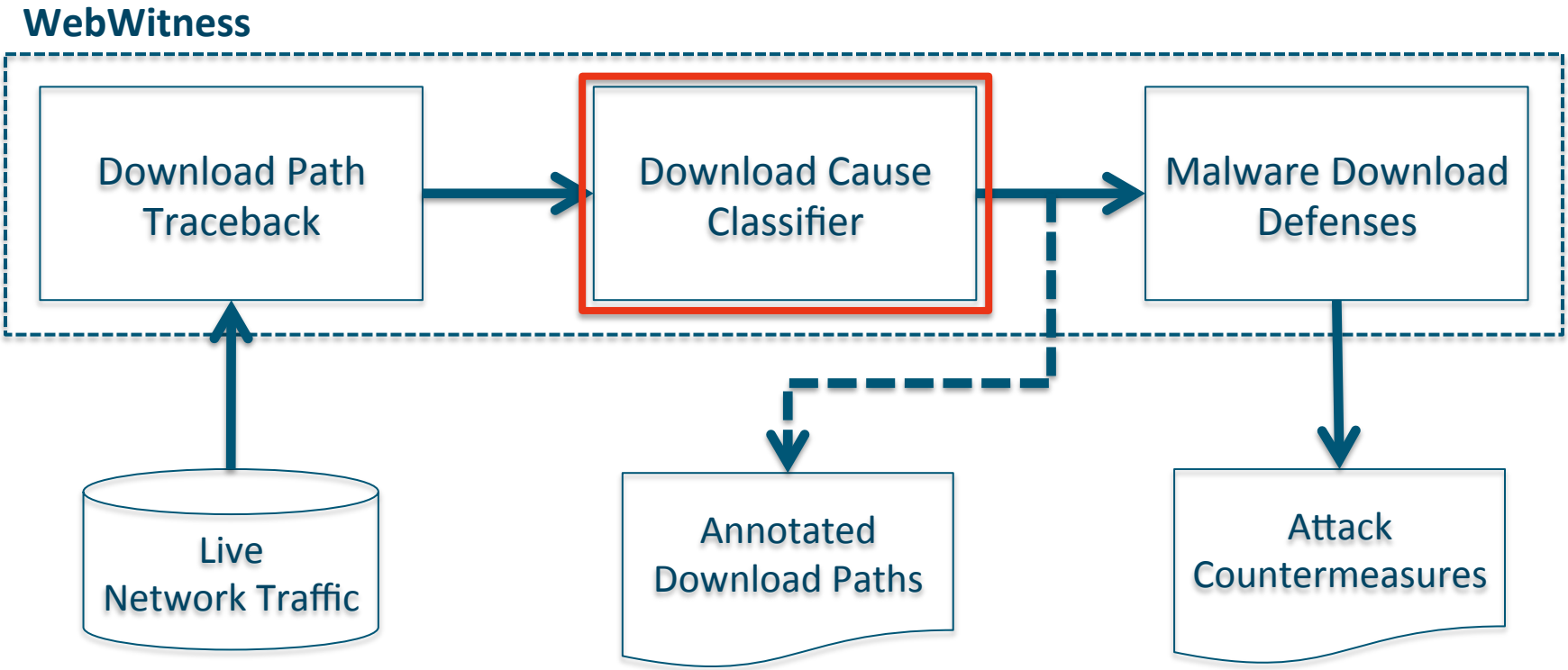
Social Eng.

53% → **95%**

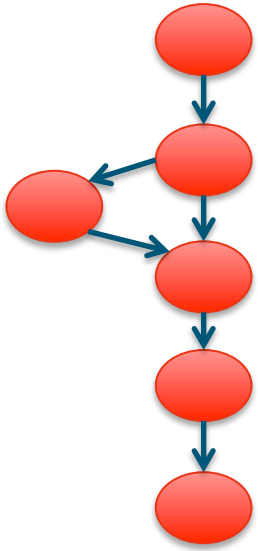
Download Path Traceback

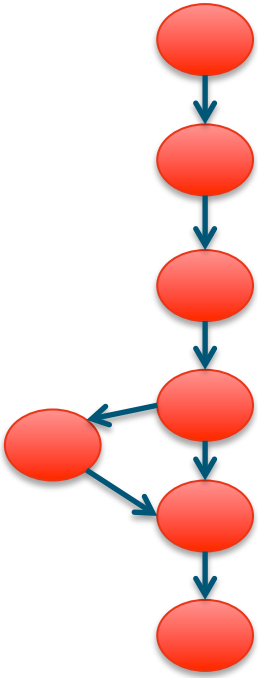


WebWitness System

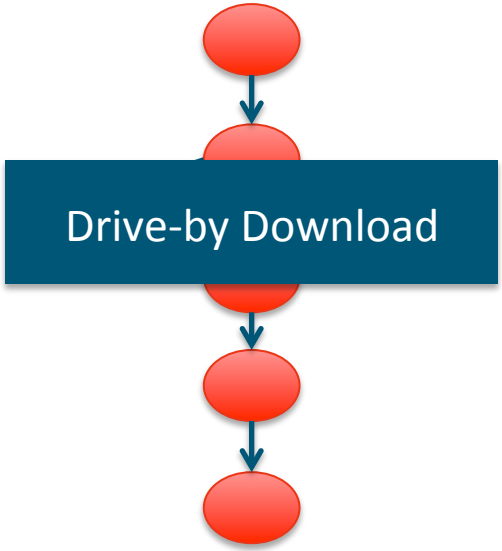


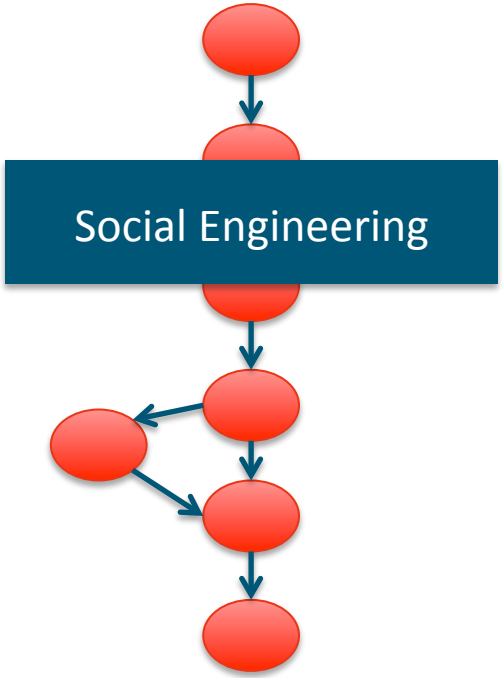
Download Paths





Download Path Classification



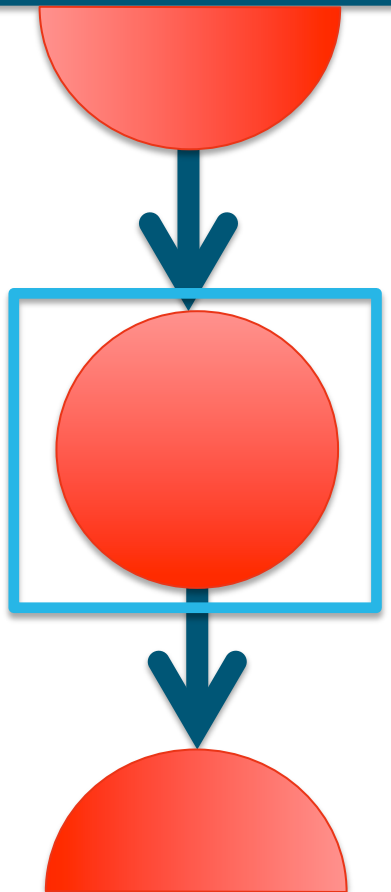


Cause Classifier Features

6 Features

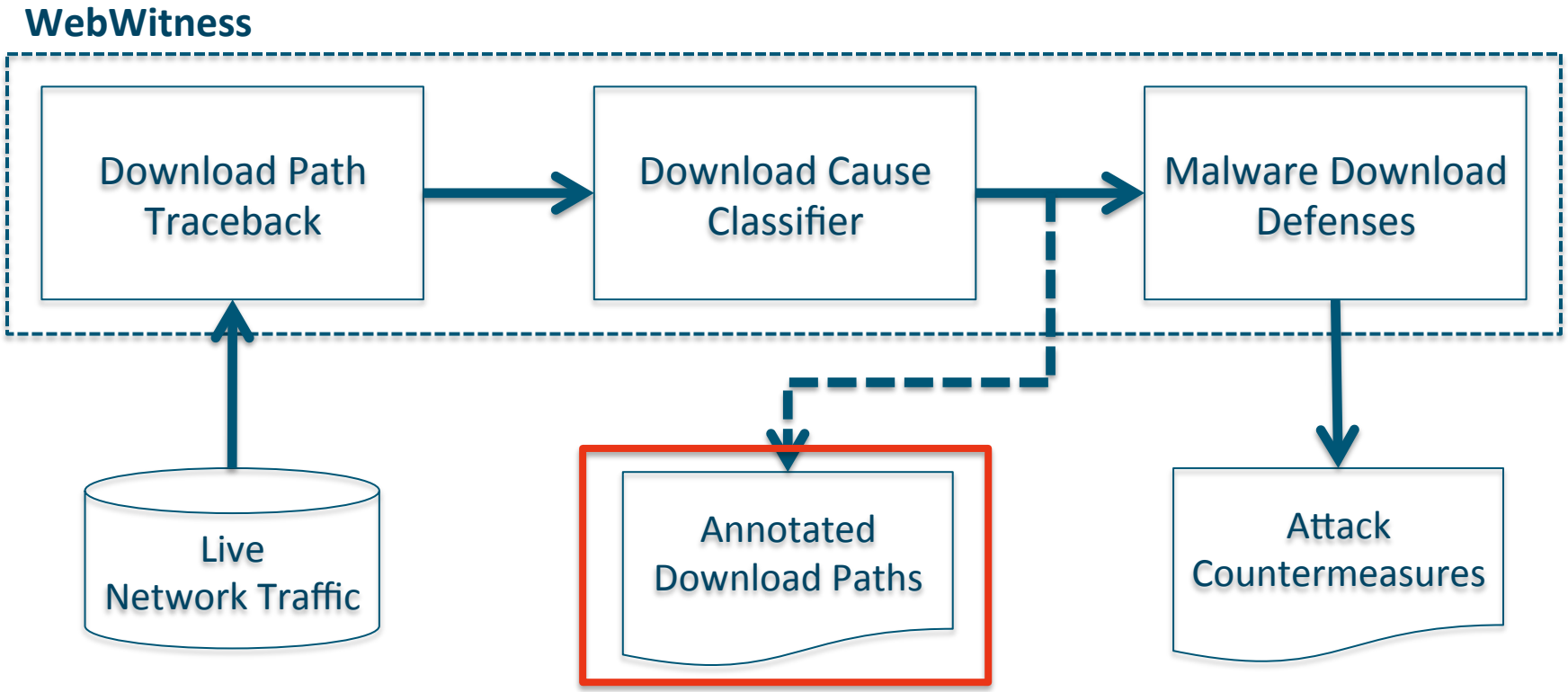
- Candidate Exploit Domain Age.
- Drive-by URL Similarity.
- Download Domain Recurrence.
- Download Referrer.
- Download Path Length.
- User-Agent Popularity.

Candidate Exploit Domain Age

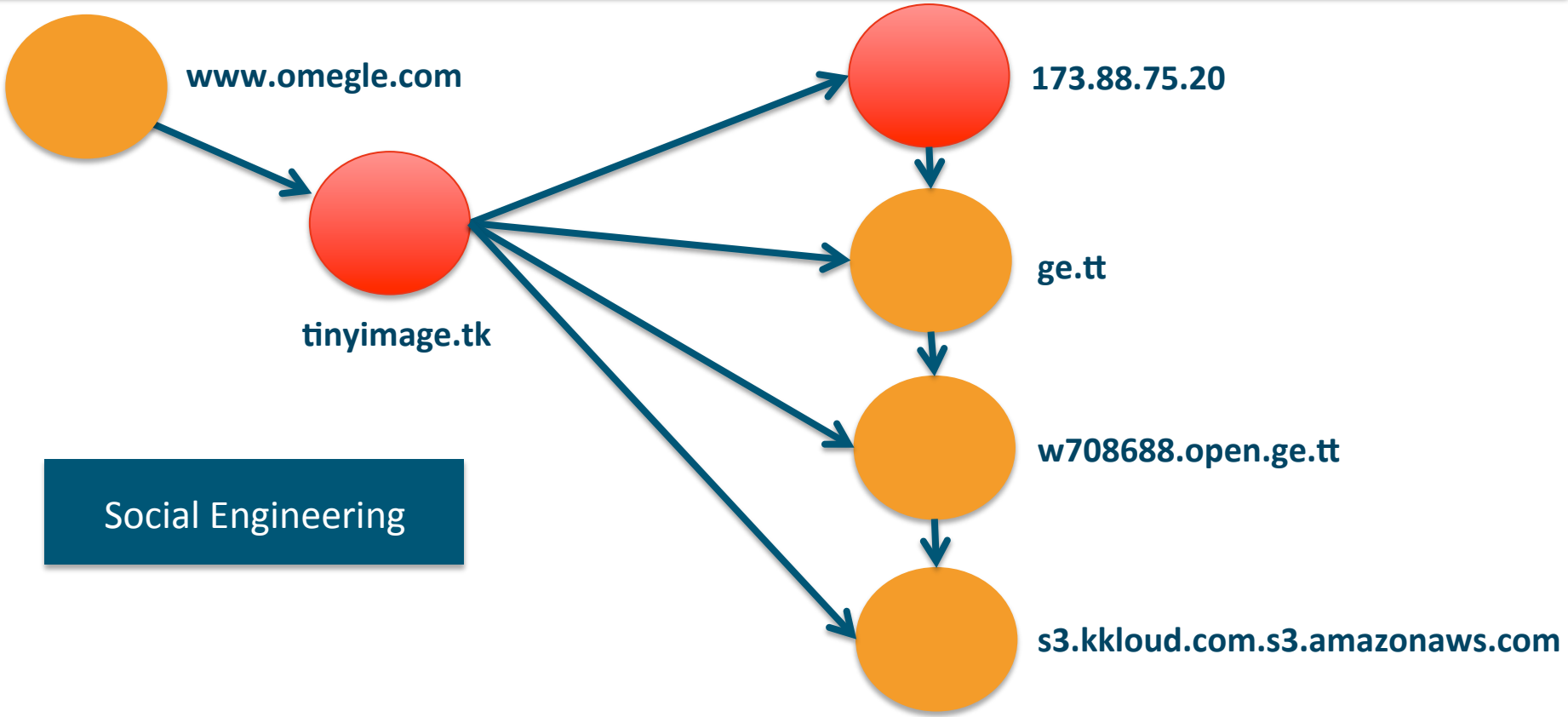


- “Commonly Exploitable” Content
  - Java
  - Flash
  - PDF
- Domain Age
  - First resolved
  - Exploit median age 0-days

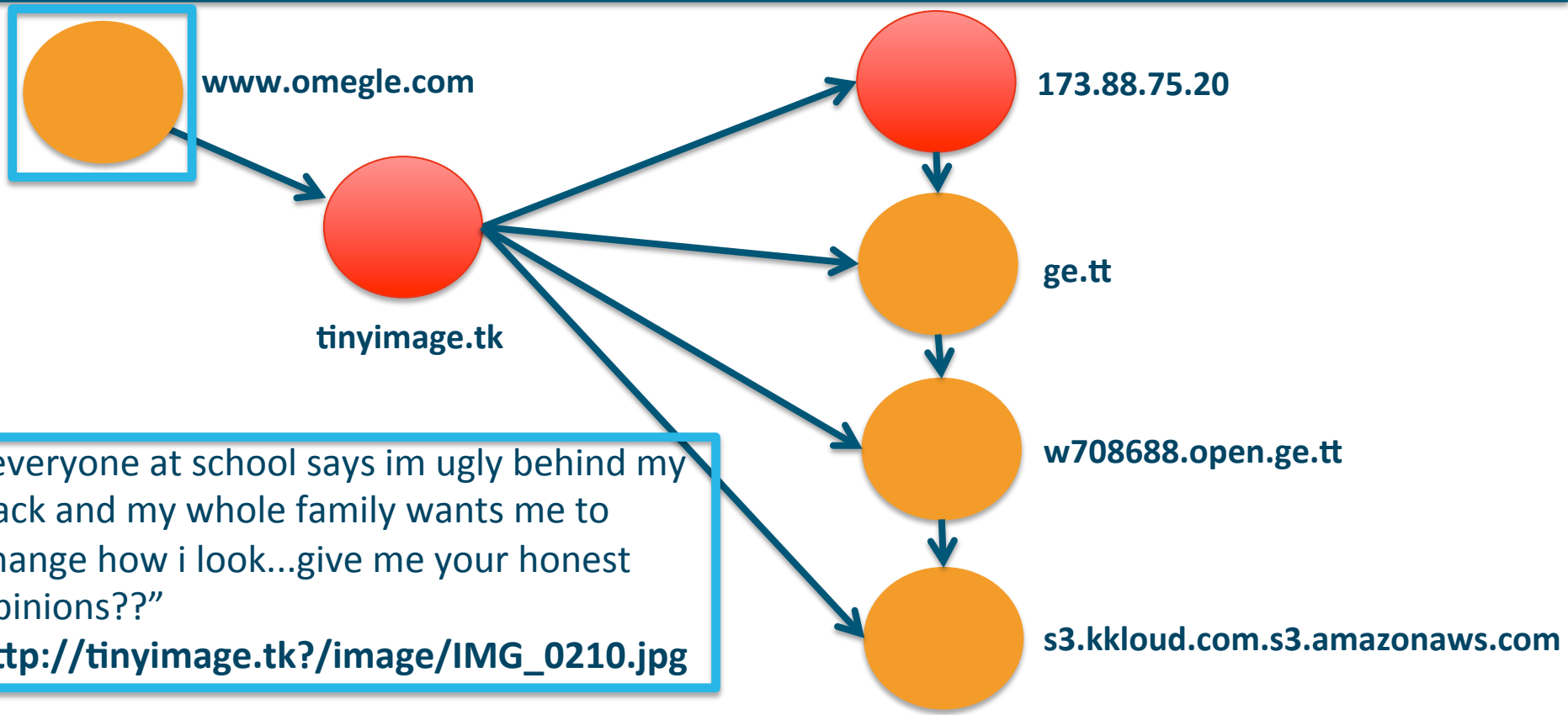
WebWitness System



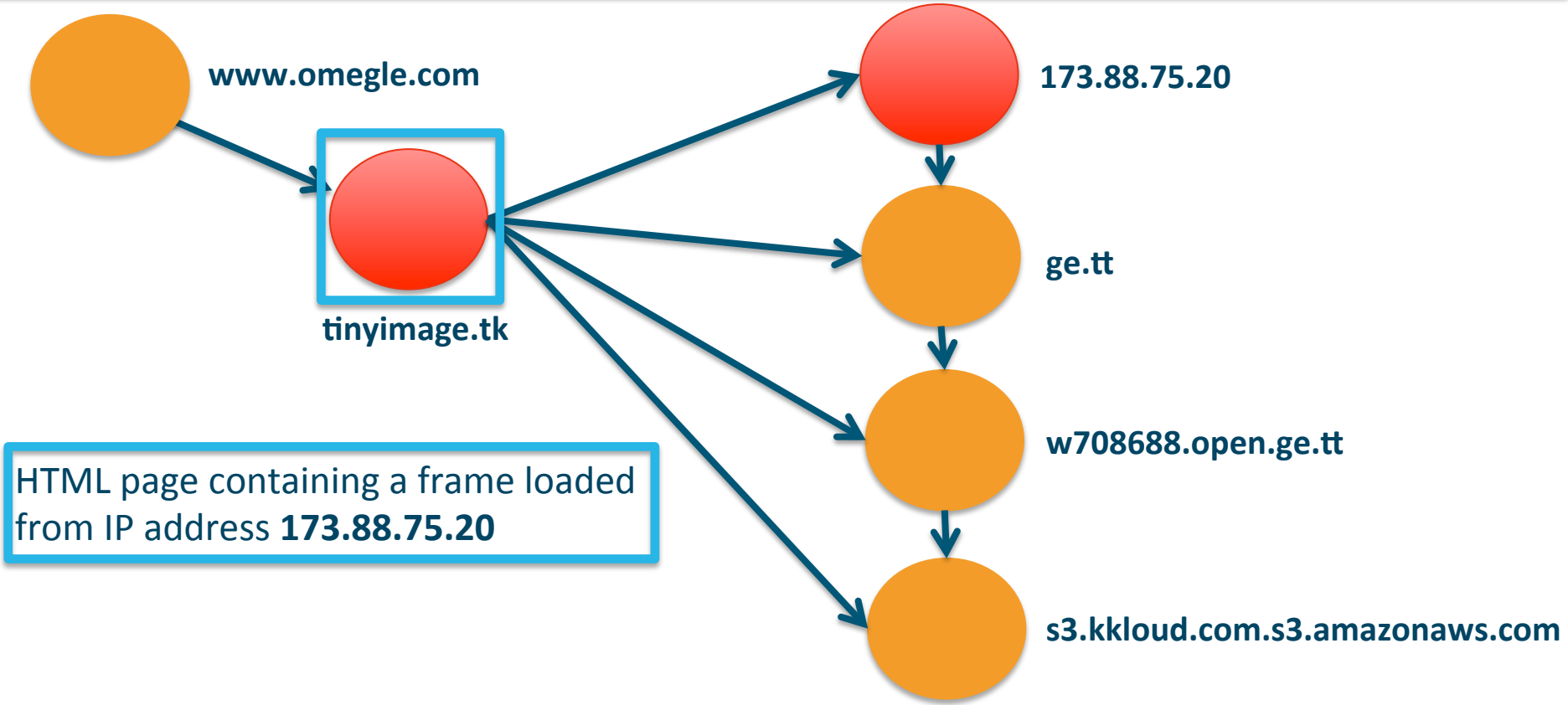
Annotated Social Engineering Path



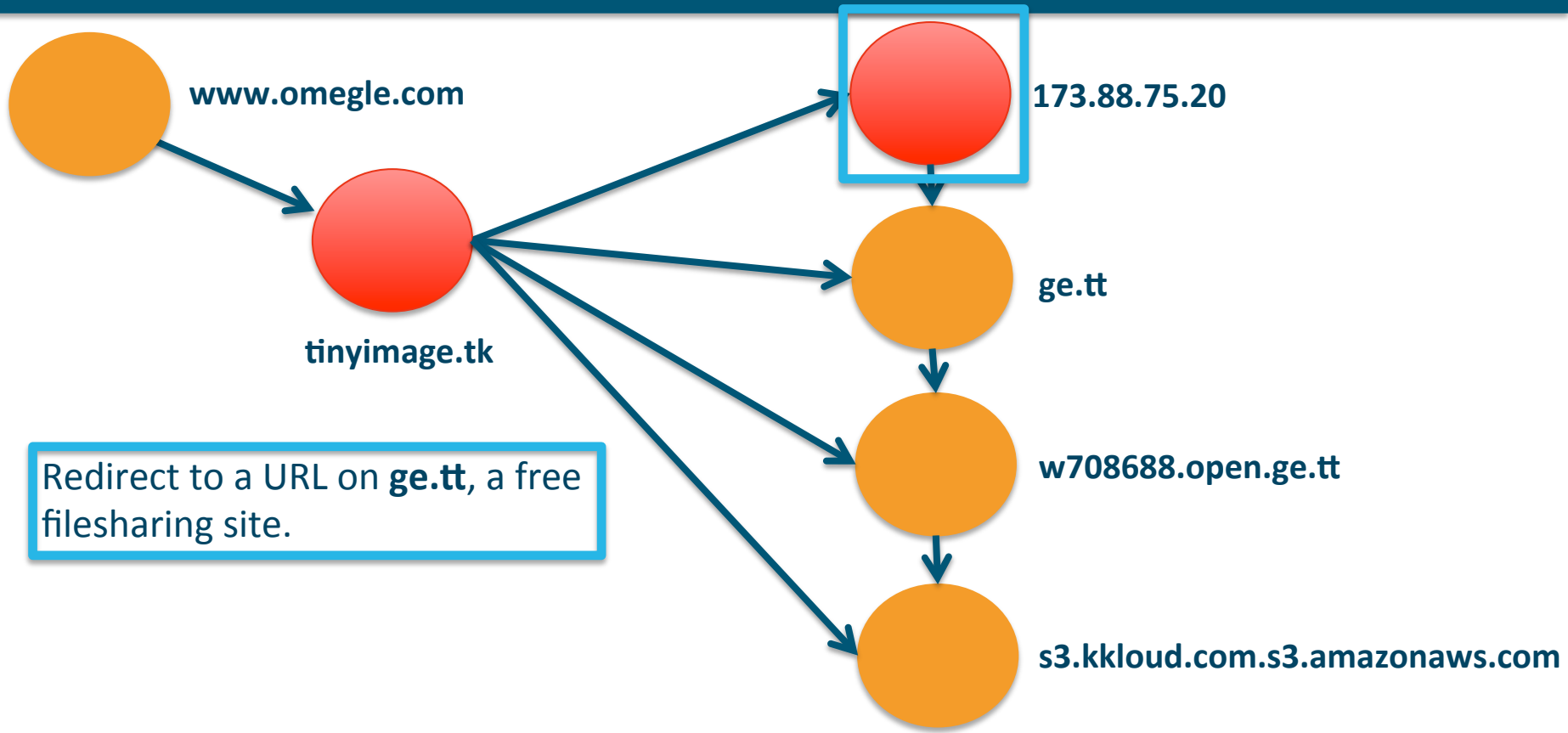
Annotated Social Engineering Path



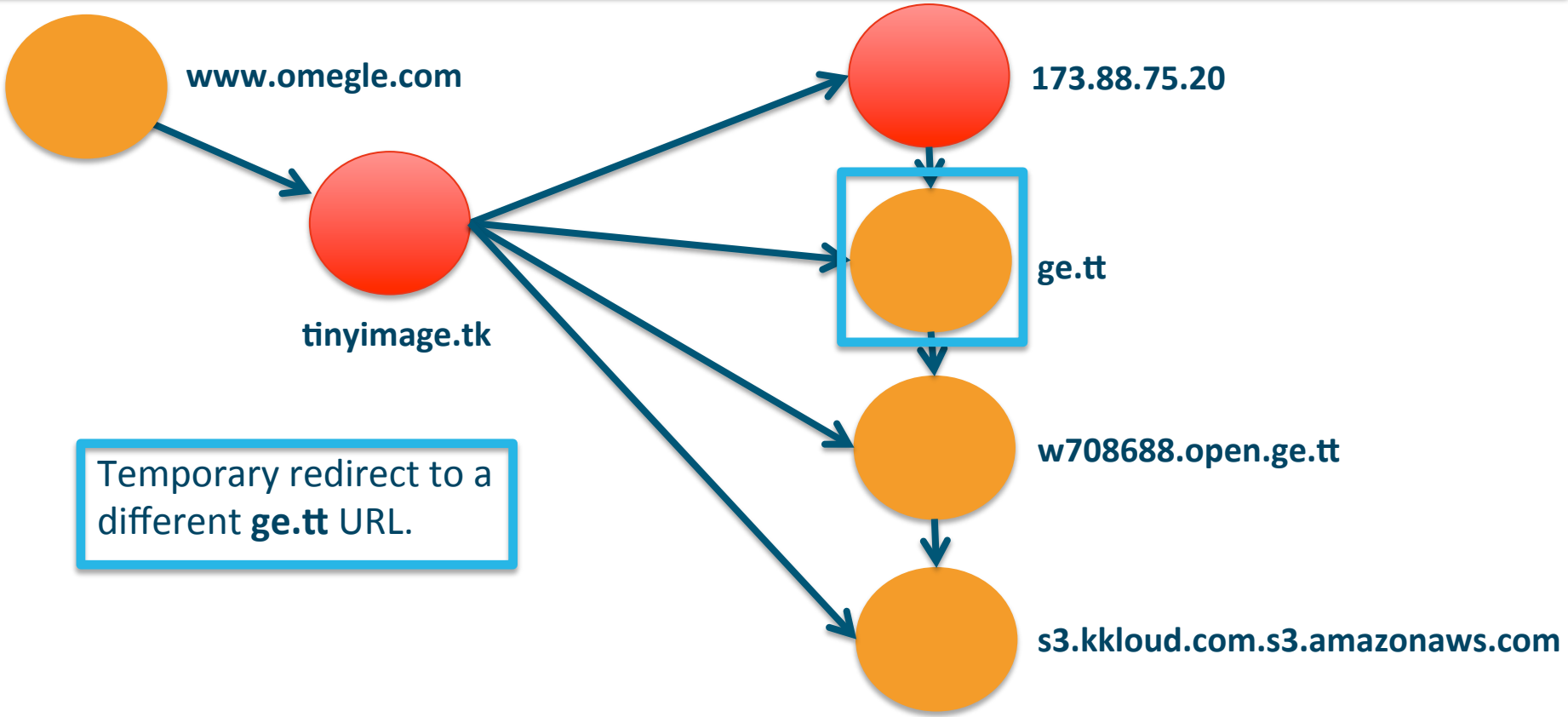
Annotated Social Engineering Path



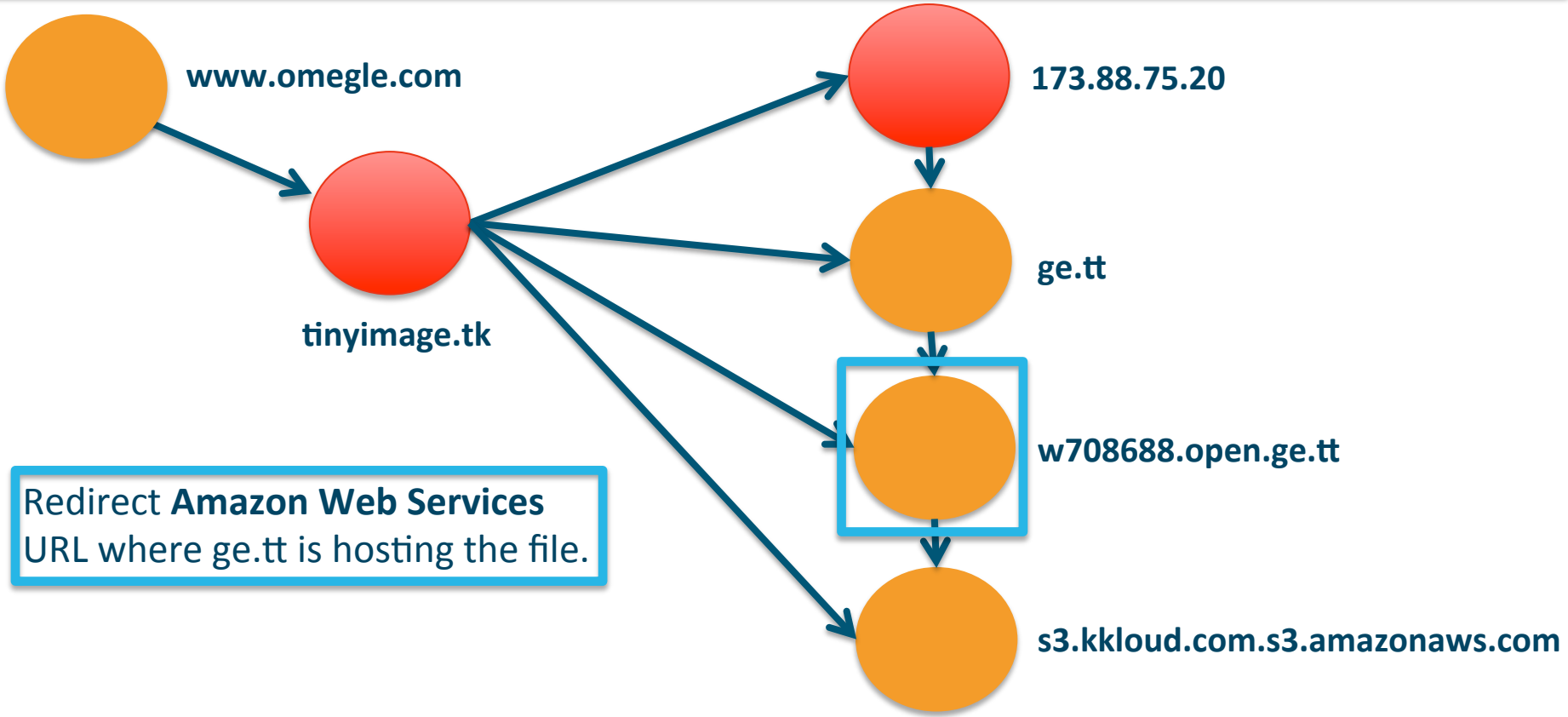
Annotated Social Engineering Path



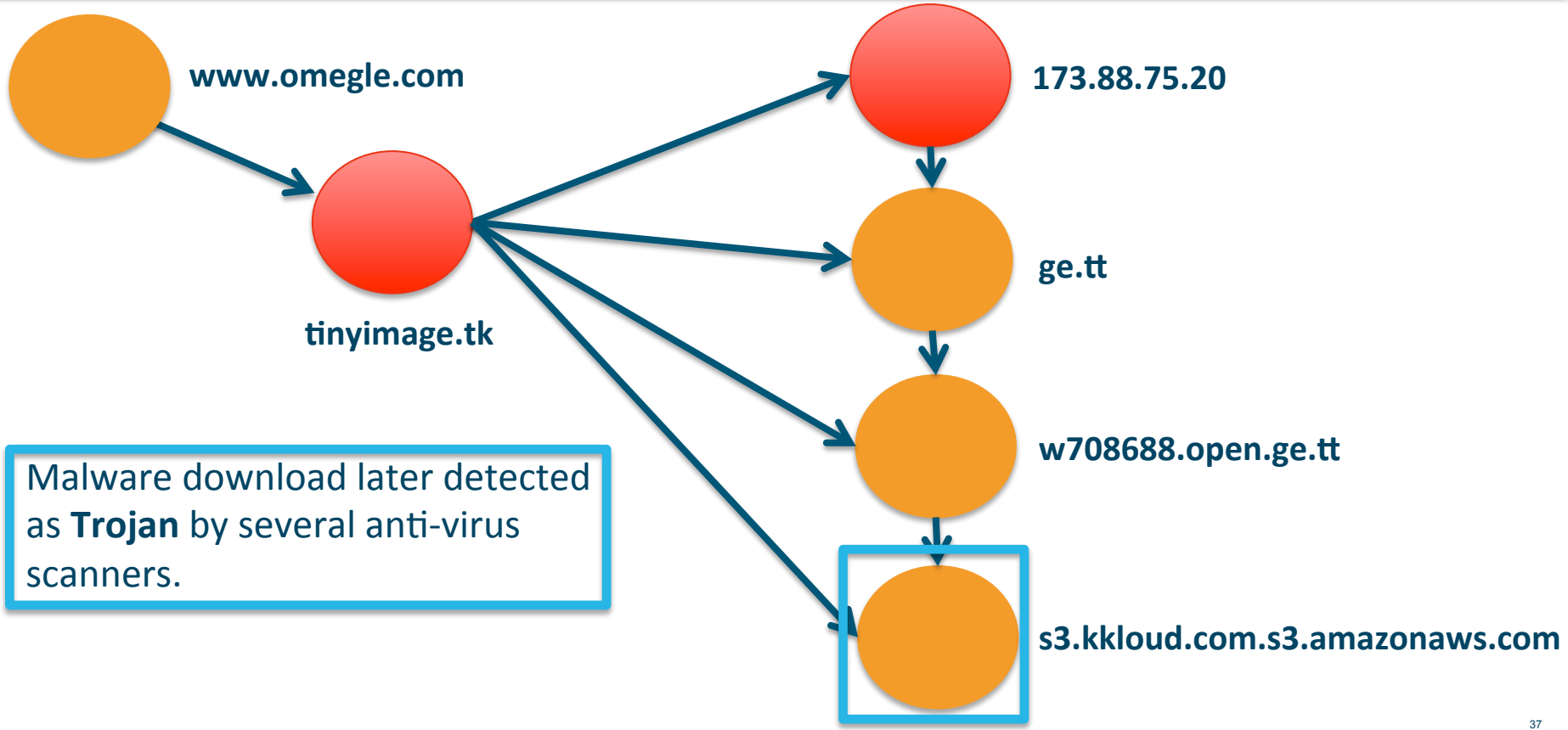
Annotated Social Engineering Path



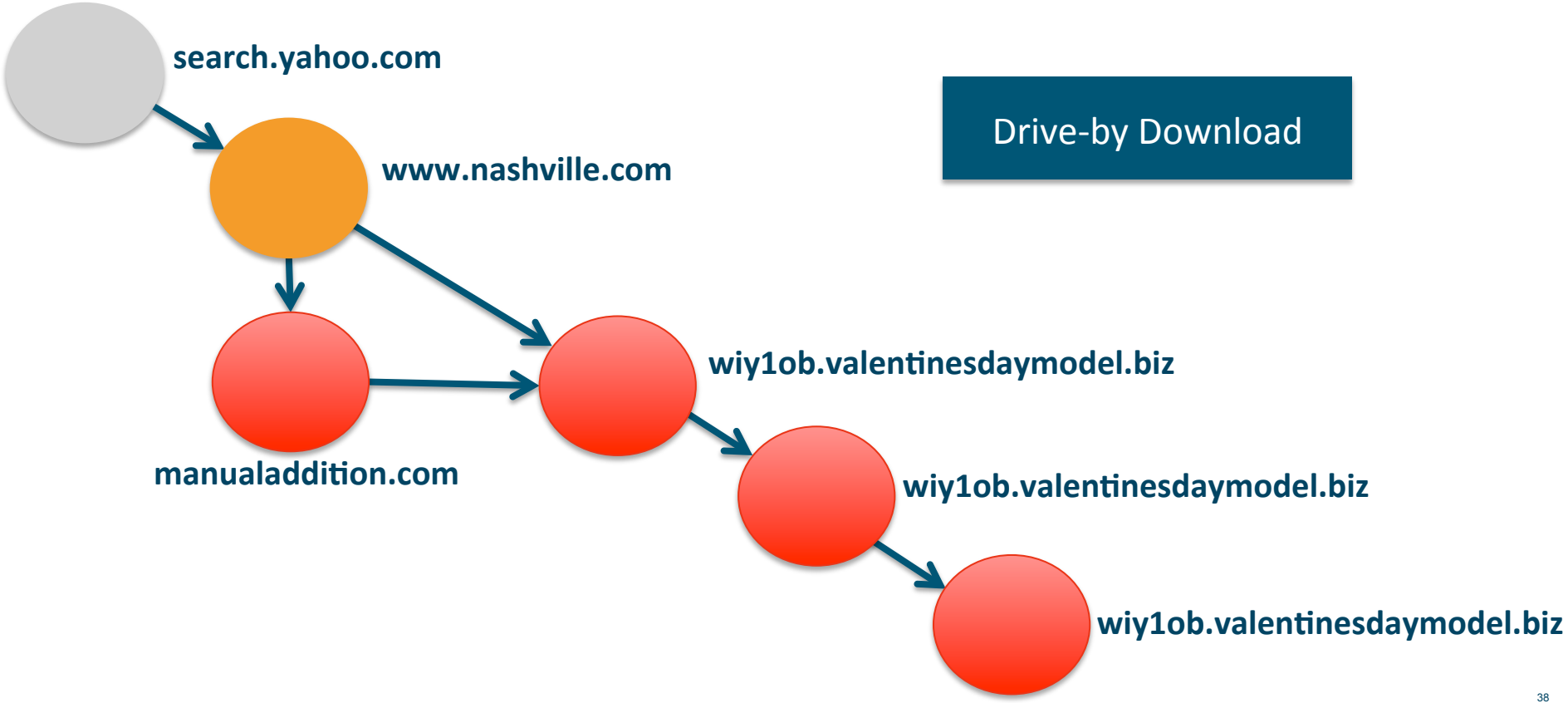
Annotated Social Engineering Path



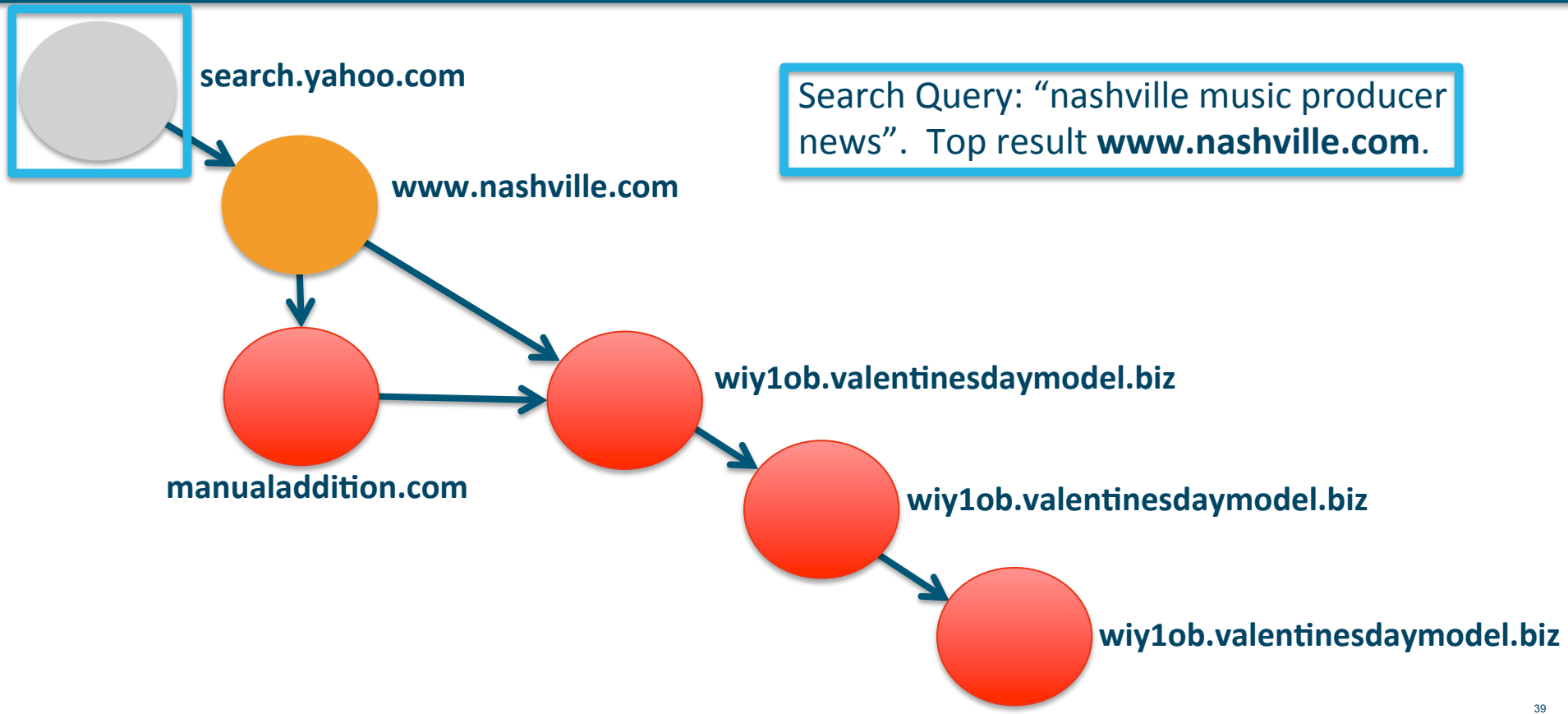
Annotated Social Engineering Path



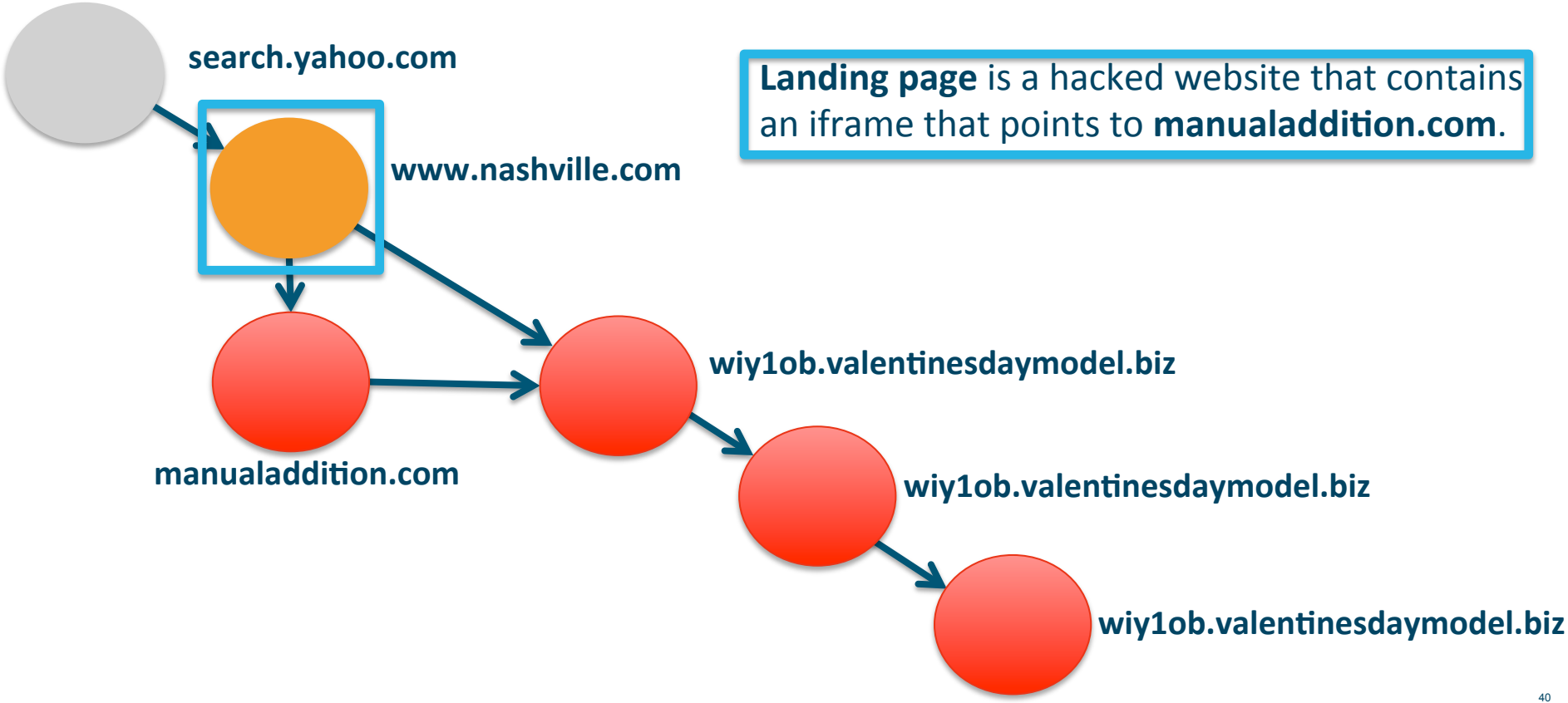
Annotated Drive-by Download Path



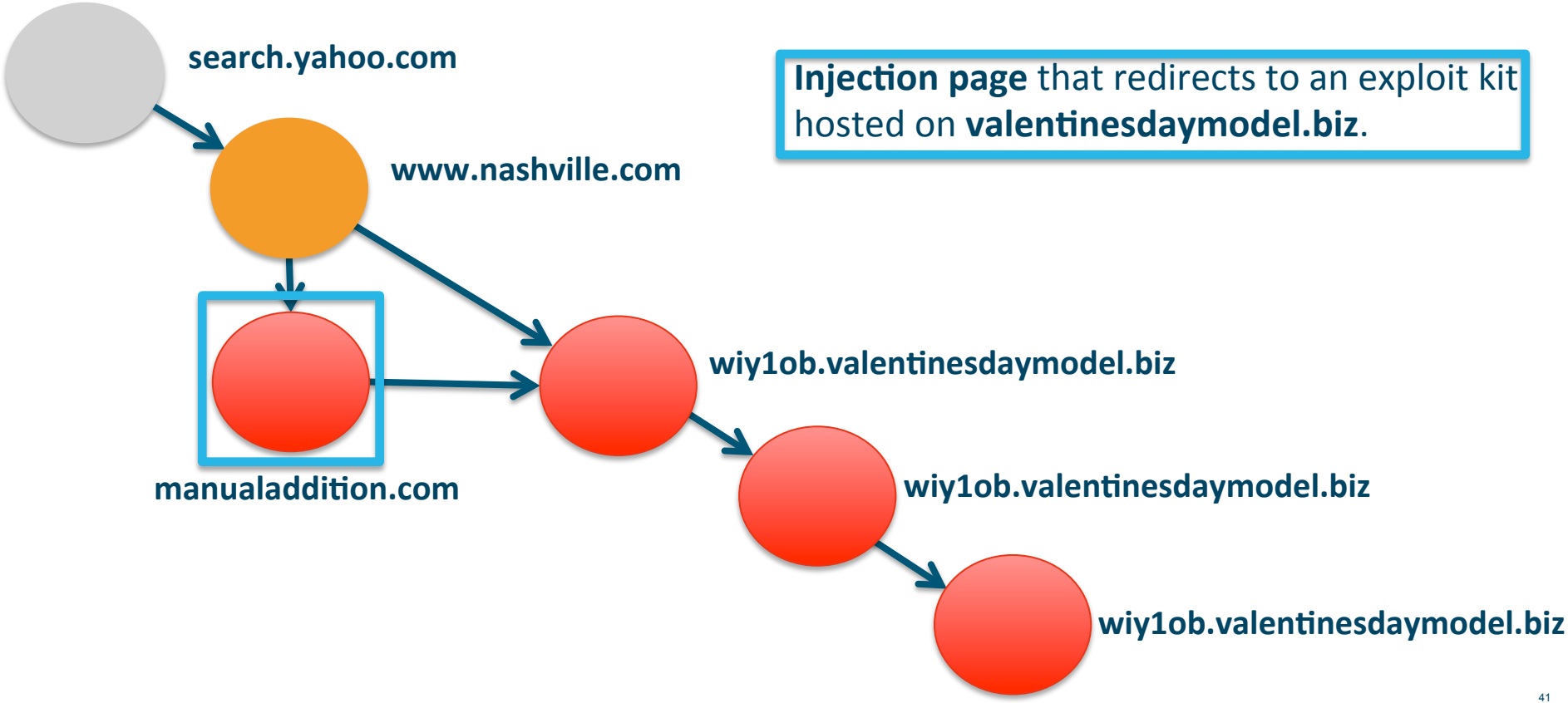
Annotated Drive-by Download Path



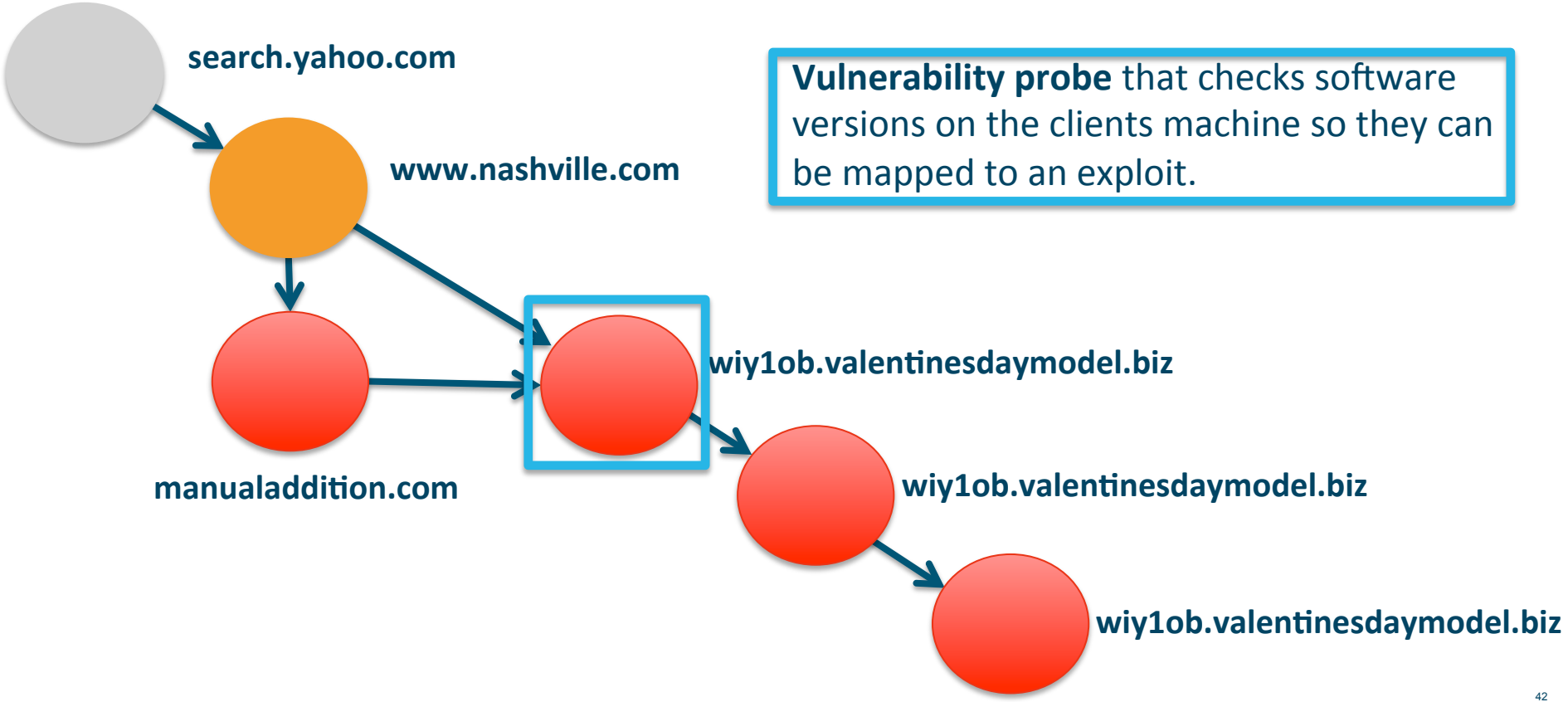
Annotated Drive-by Download Path



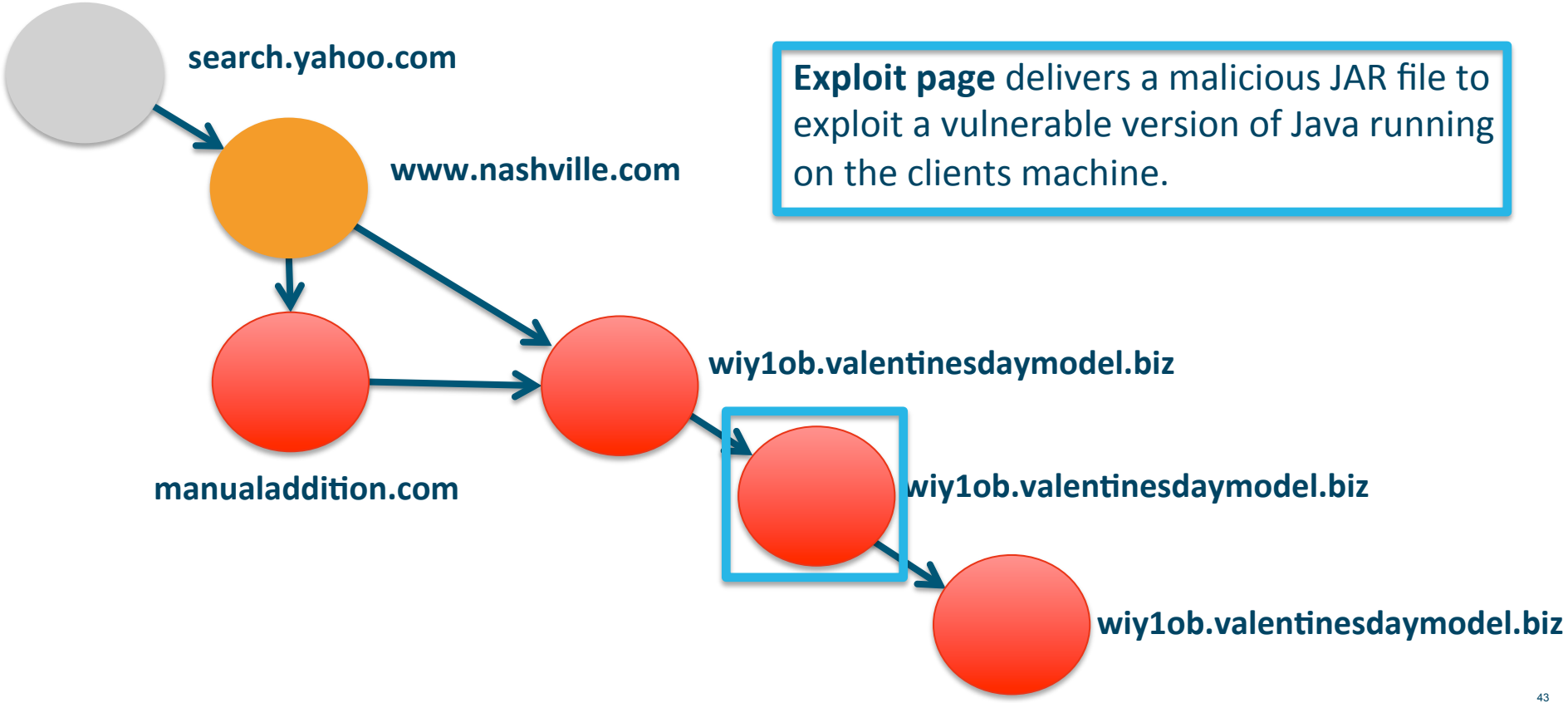
Annotated Drive-by Download Path



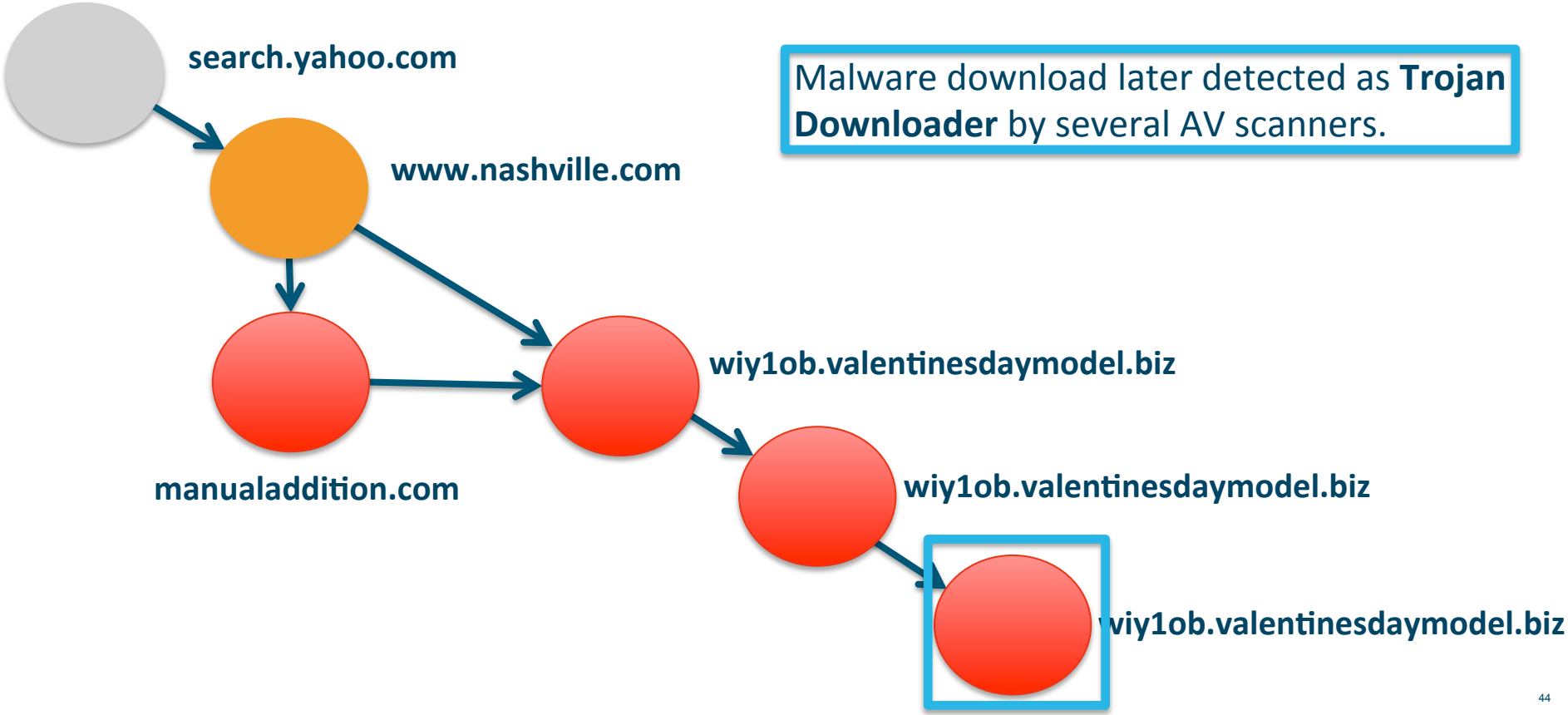
Annotated Drive-by Download Path



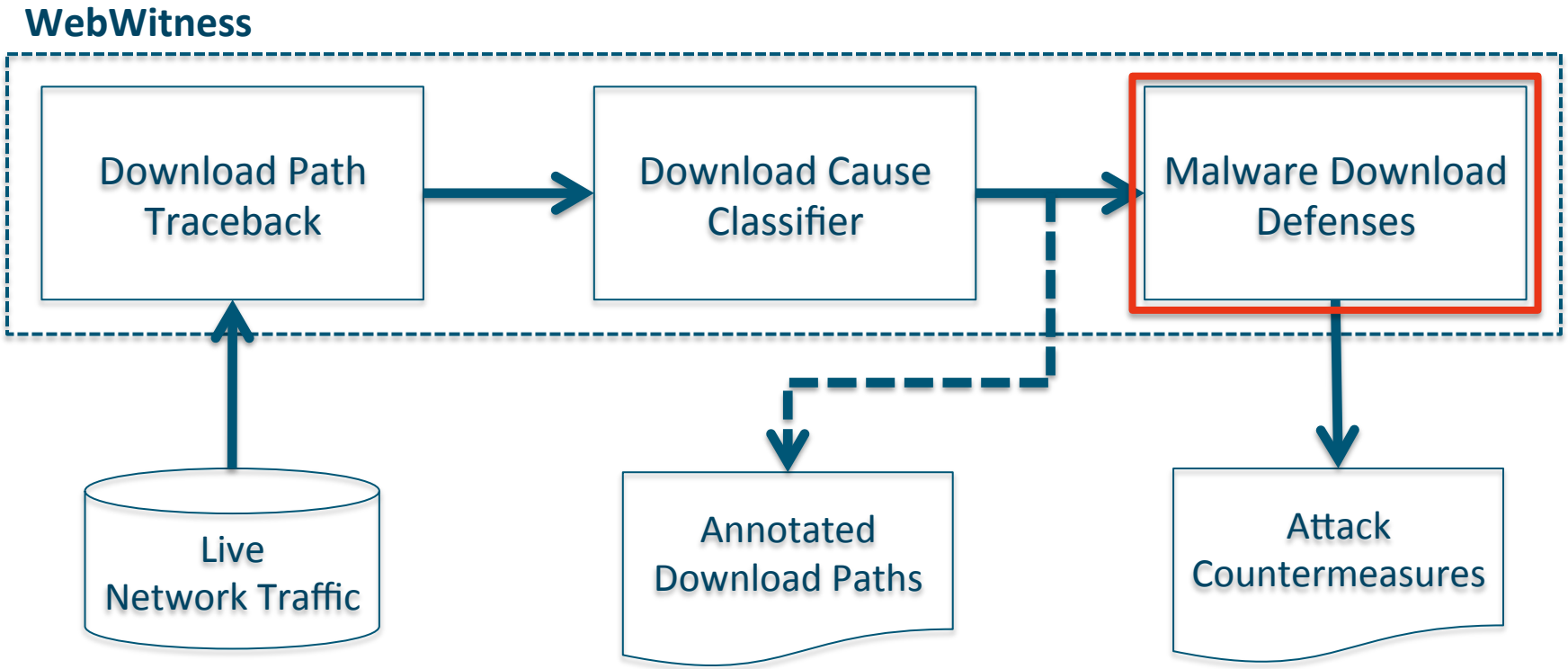
Annotated Drive-by Download Path



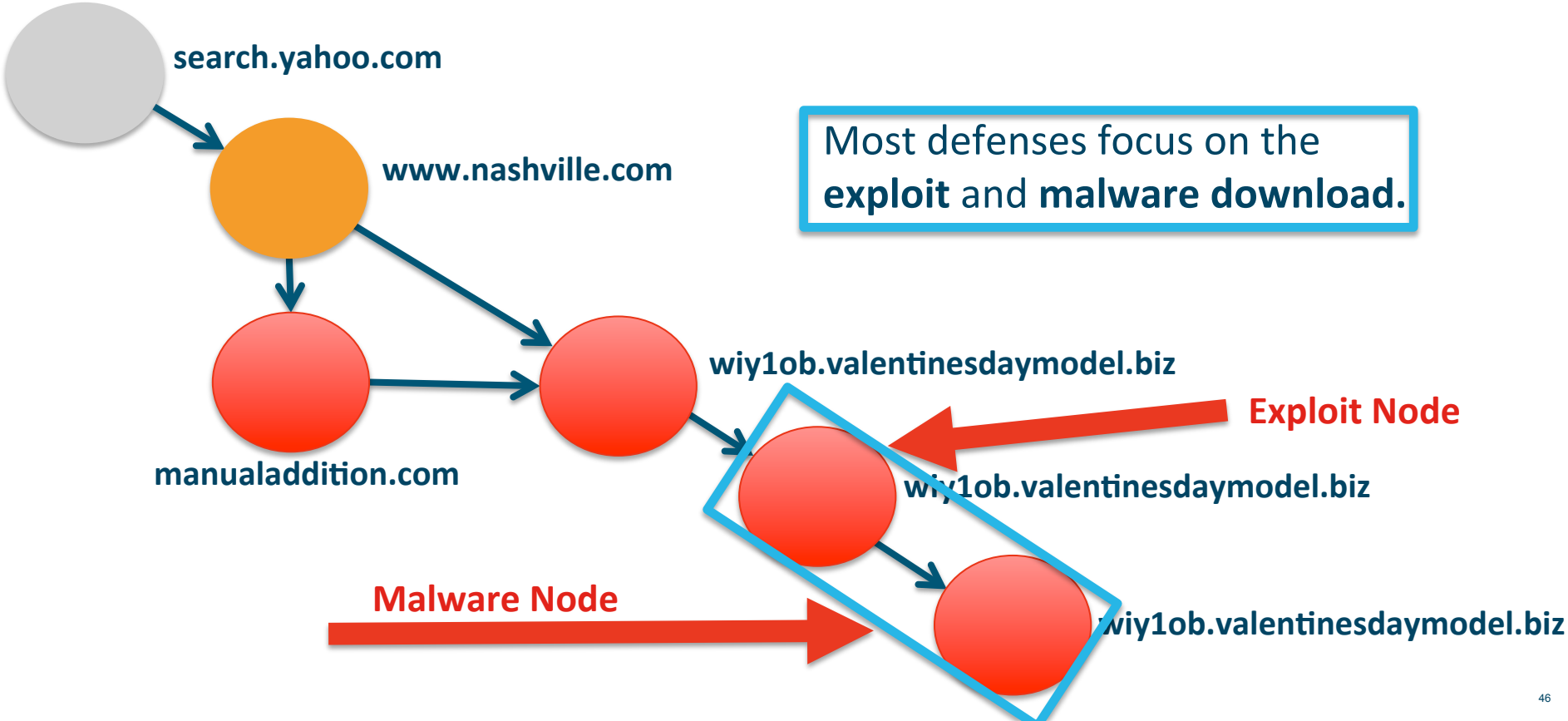
Annotated Drive-by Download Path



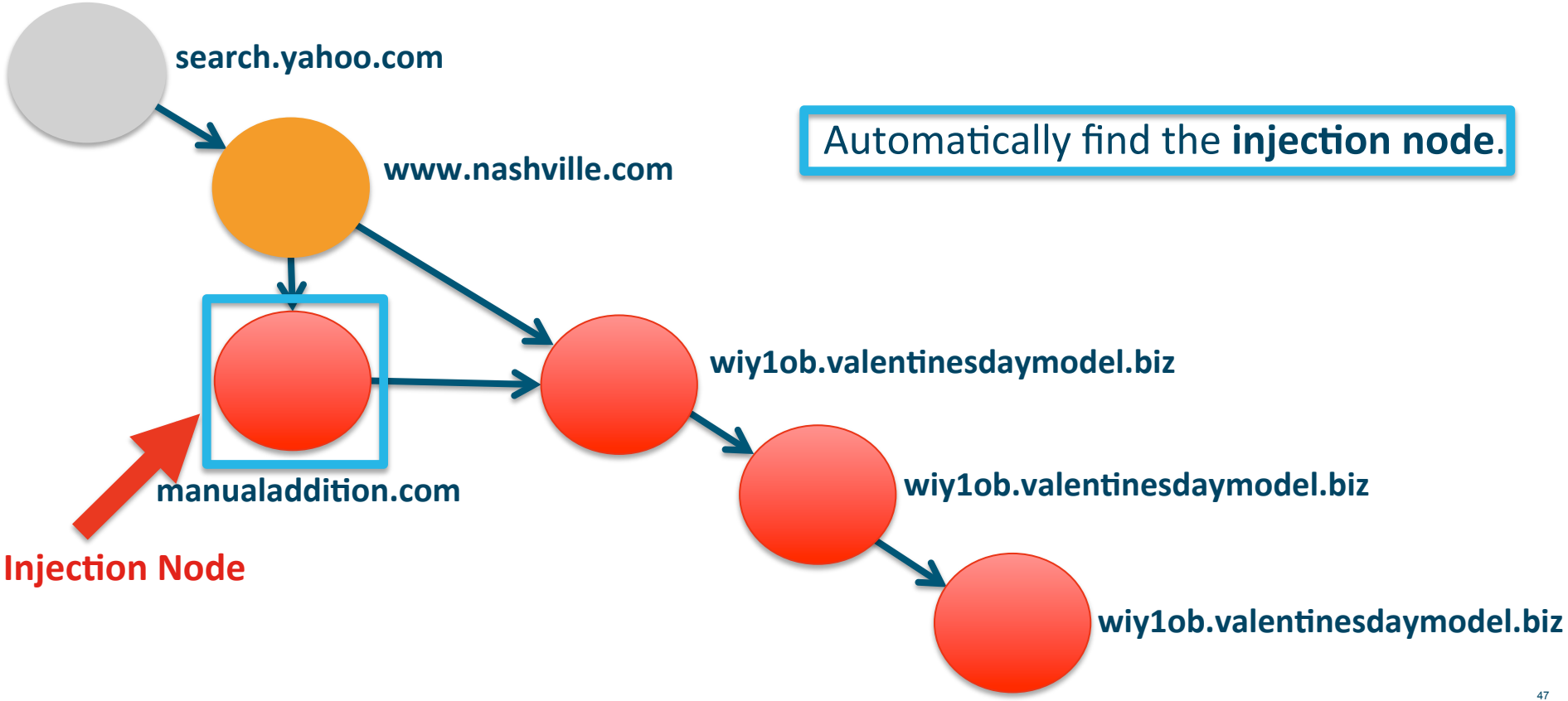
WebWitness System



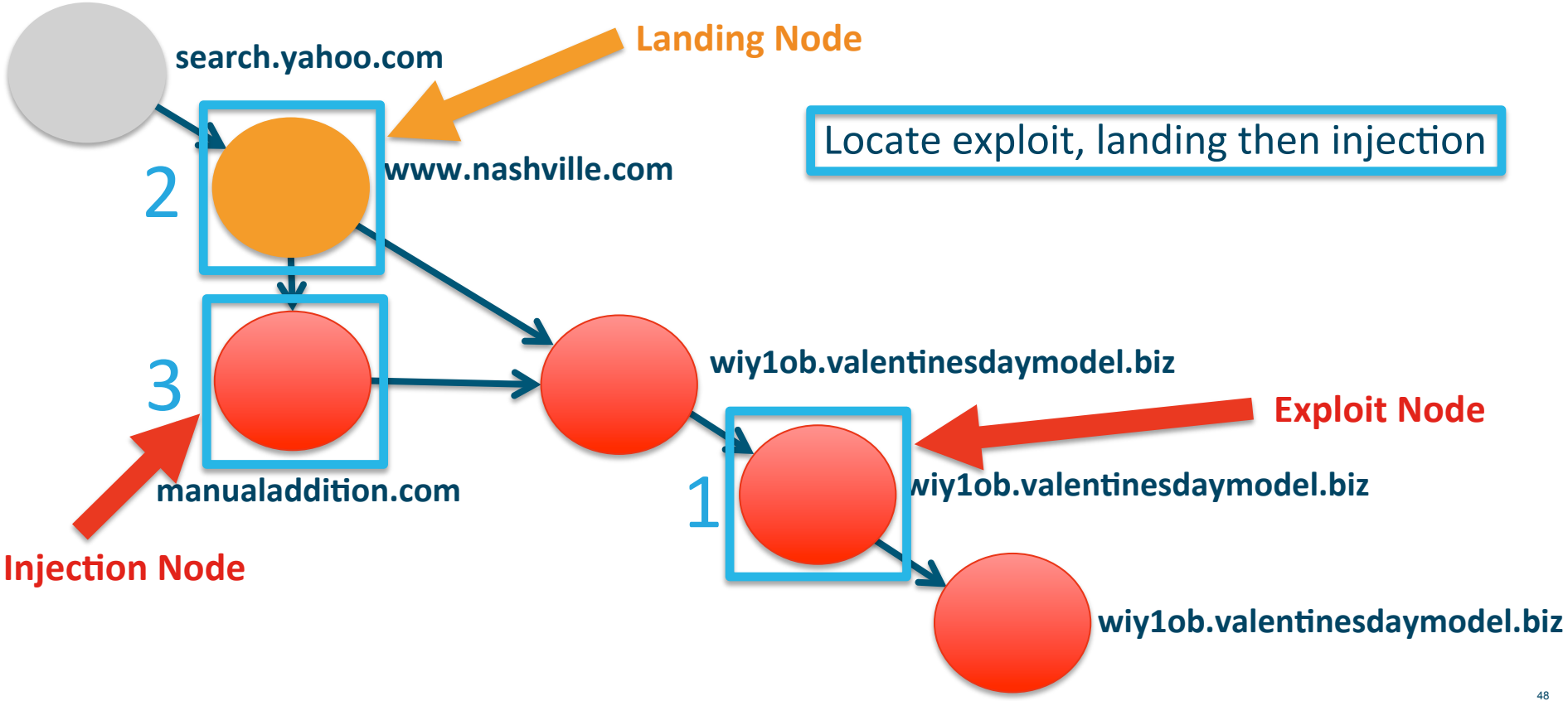
Drive-by Download Defense



Drive-by Download Defense

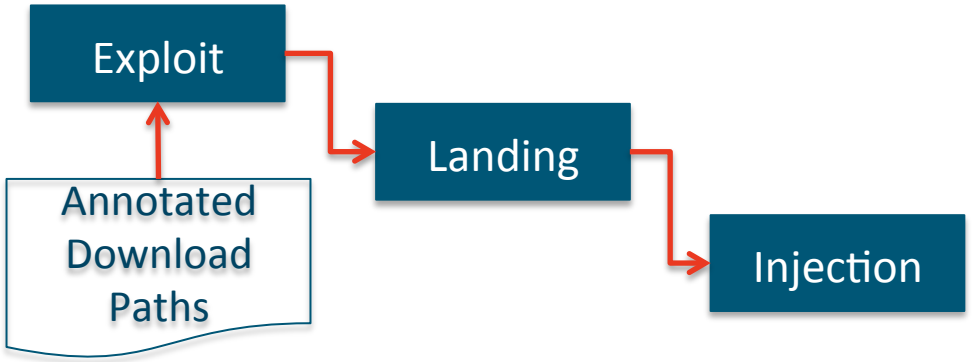


Drive-by Download Defense

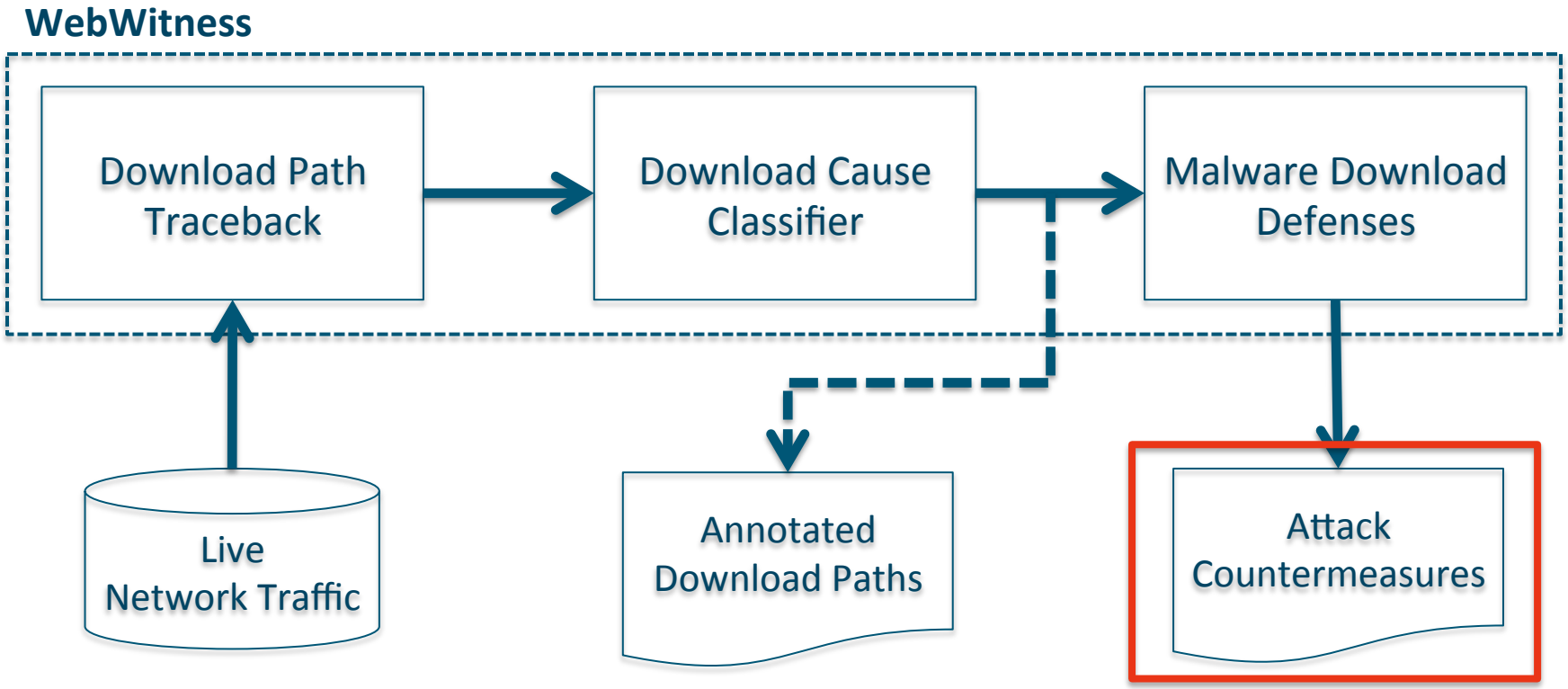


Node Classification Features

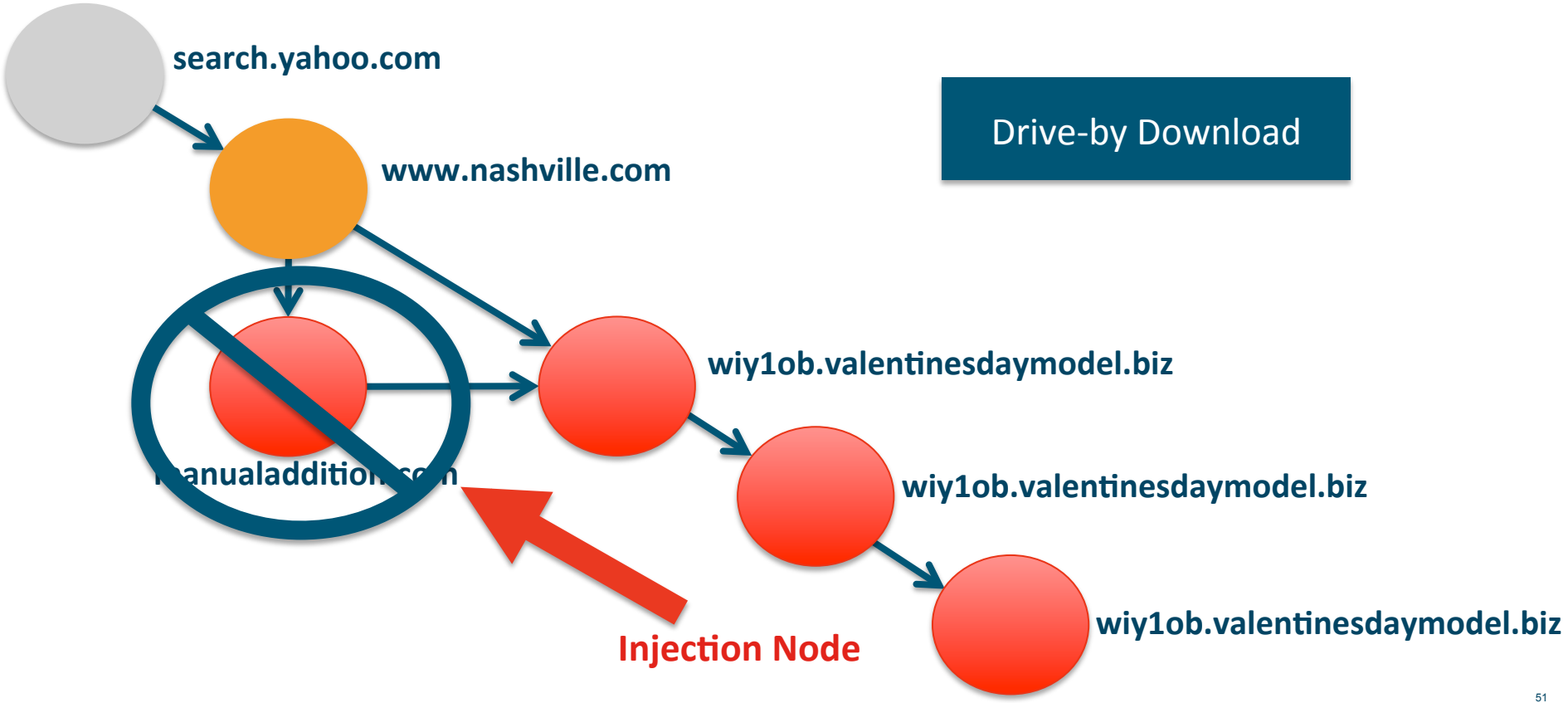
- **Exploit**
  - Hops to download page.
  - Commonly exploitable content.
  - Domain age.
  - Same domain.
- **Landing**
  - Hops to exploit page.
  - Domain age.
  - Same domain.
- **Injection**
  - On path.
  - Advertisement.
  - Domain age.
  - Successors.
  - Same domain.



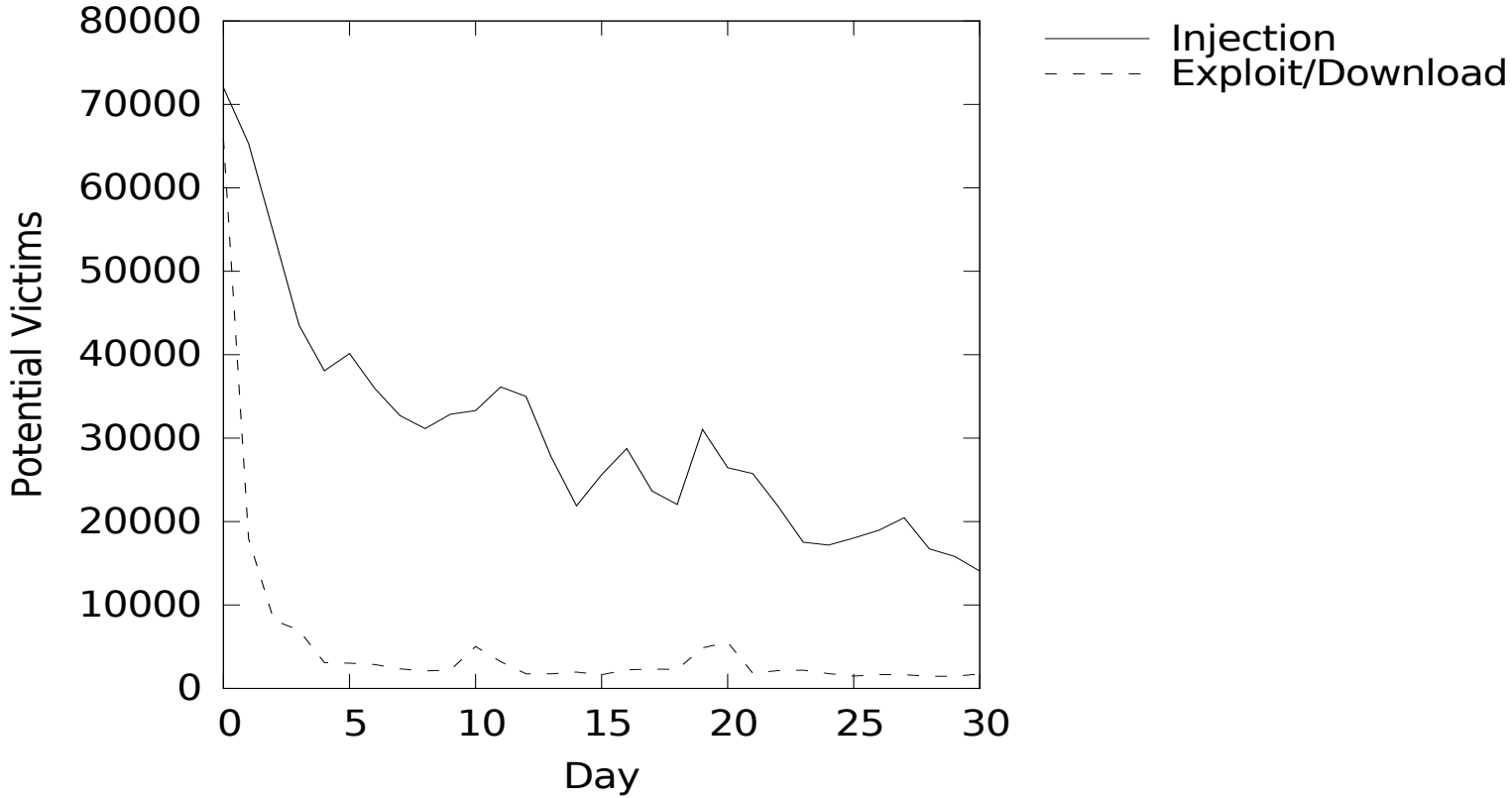
WebWitness System



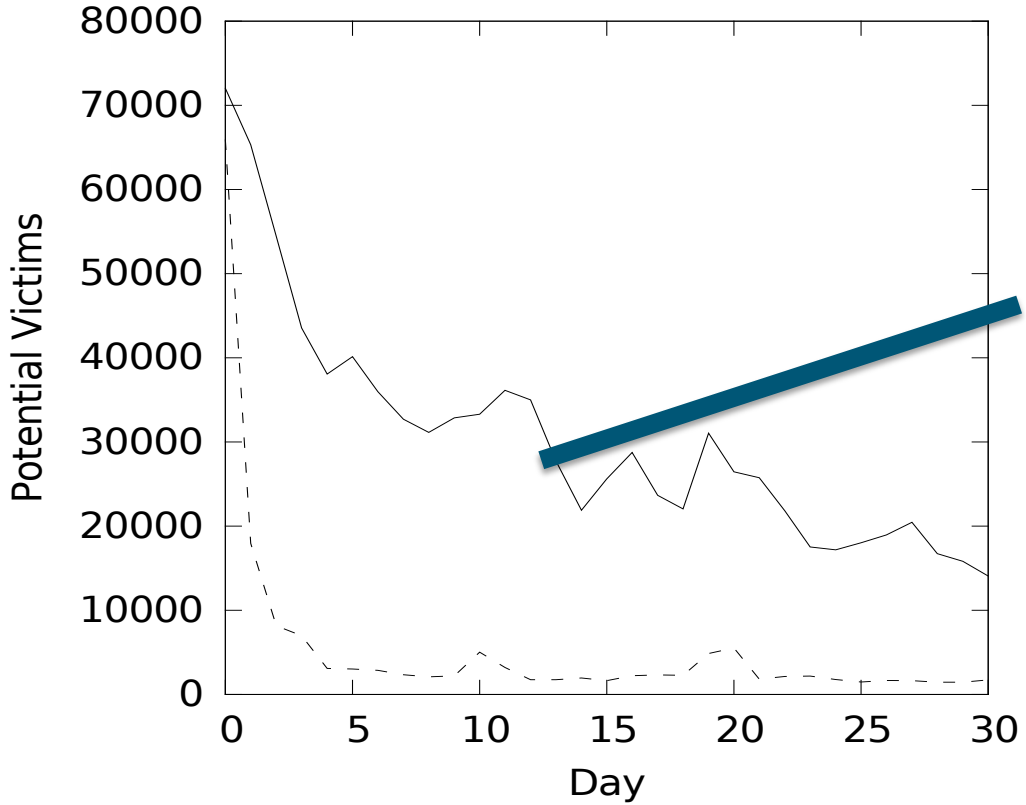
Block Injection Domain



Saved Victims – Blocking Injection Domain

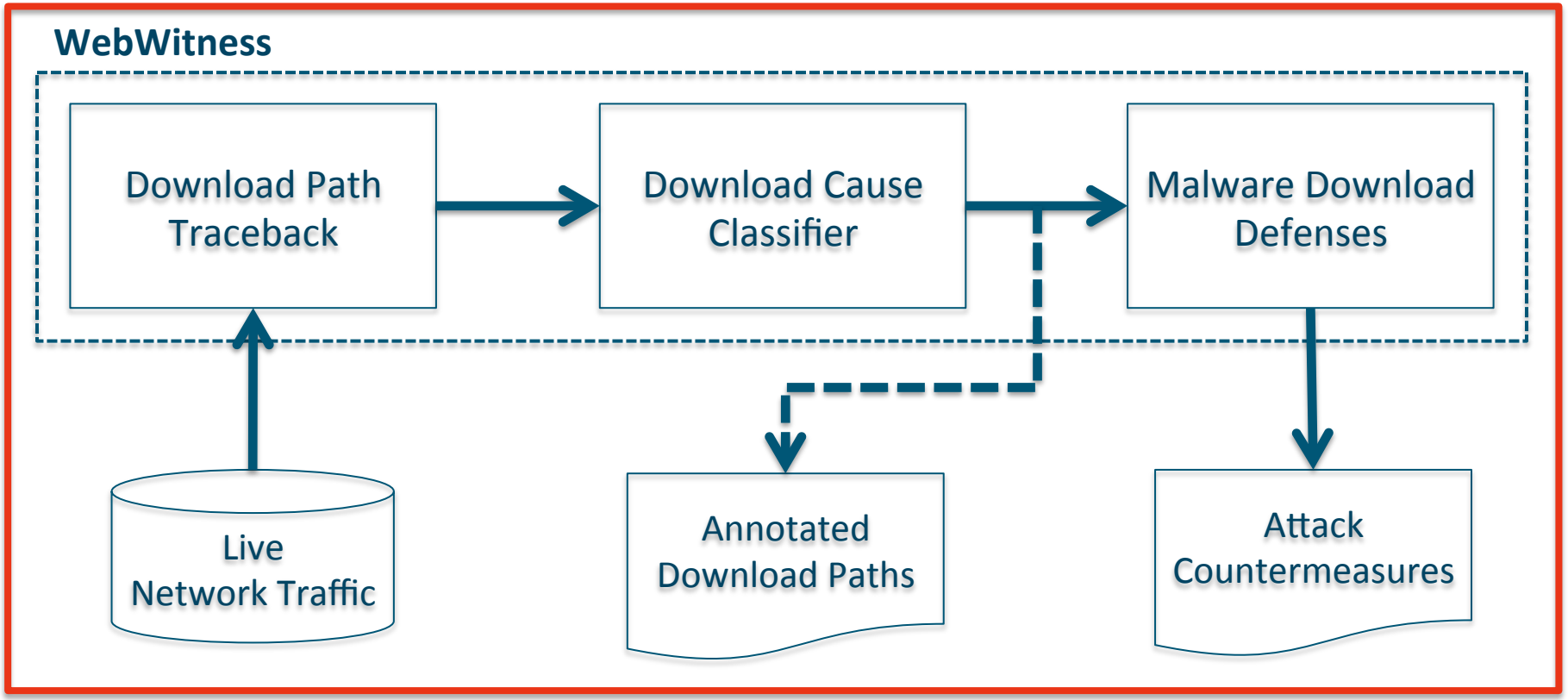


Saved Victims – Blocking Injection Domain



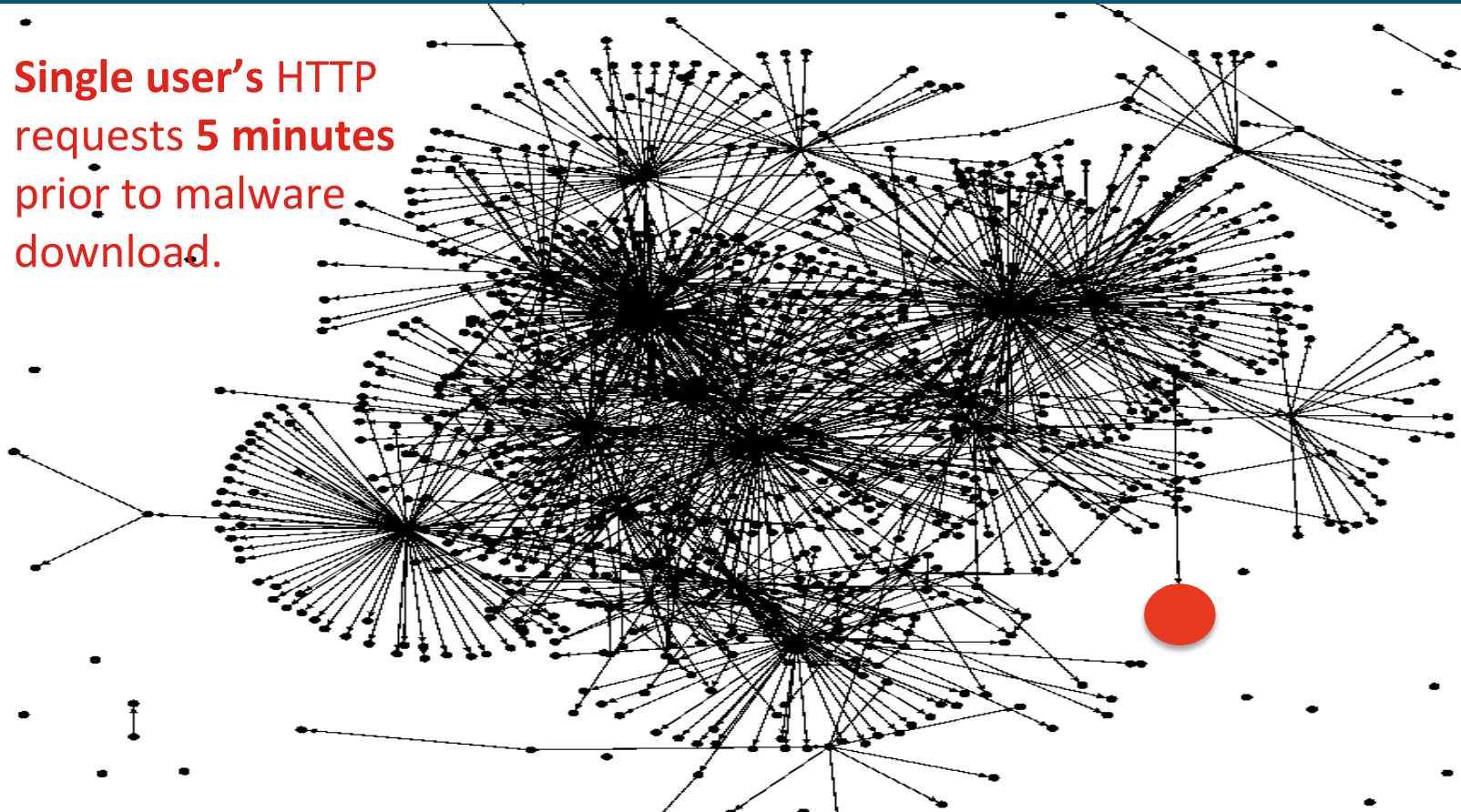
6 *times* more saved victims!

Conclusion

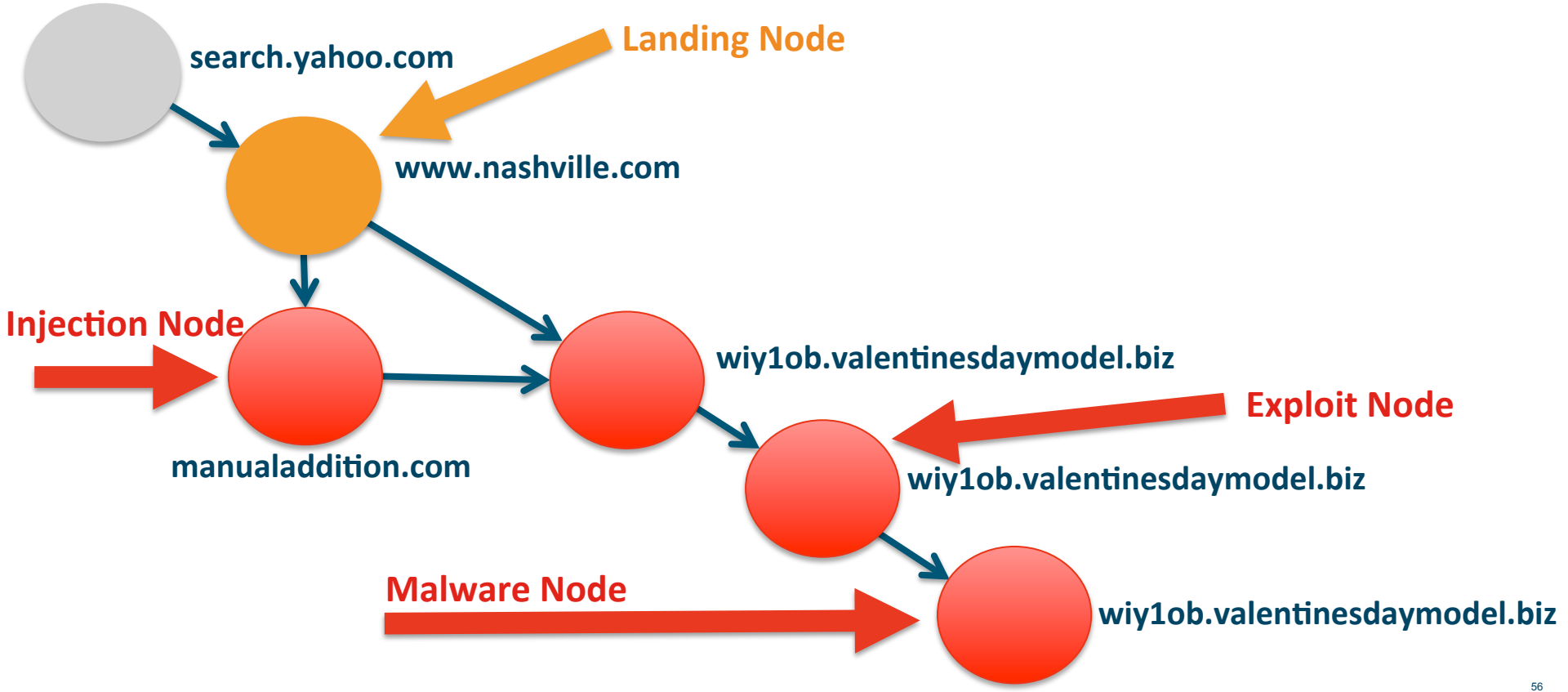

WebWitness
Download Path Traceback
Download Cause Classifier
Malware Download Defenses
Live Network Traffic
Annotated Download Paths
Attack Countermeasures

Web Traffic Graph

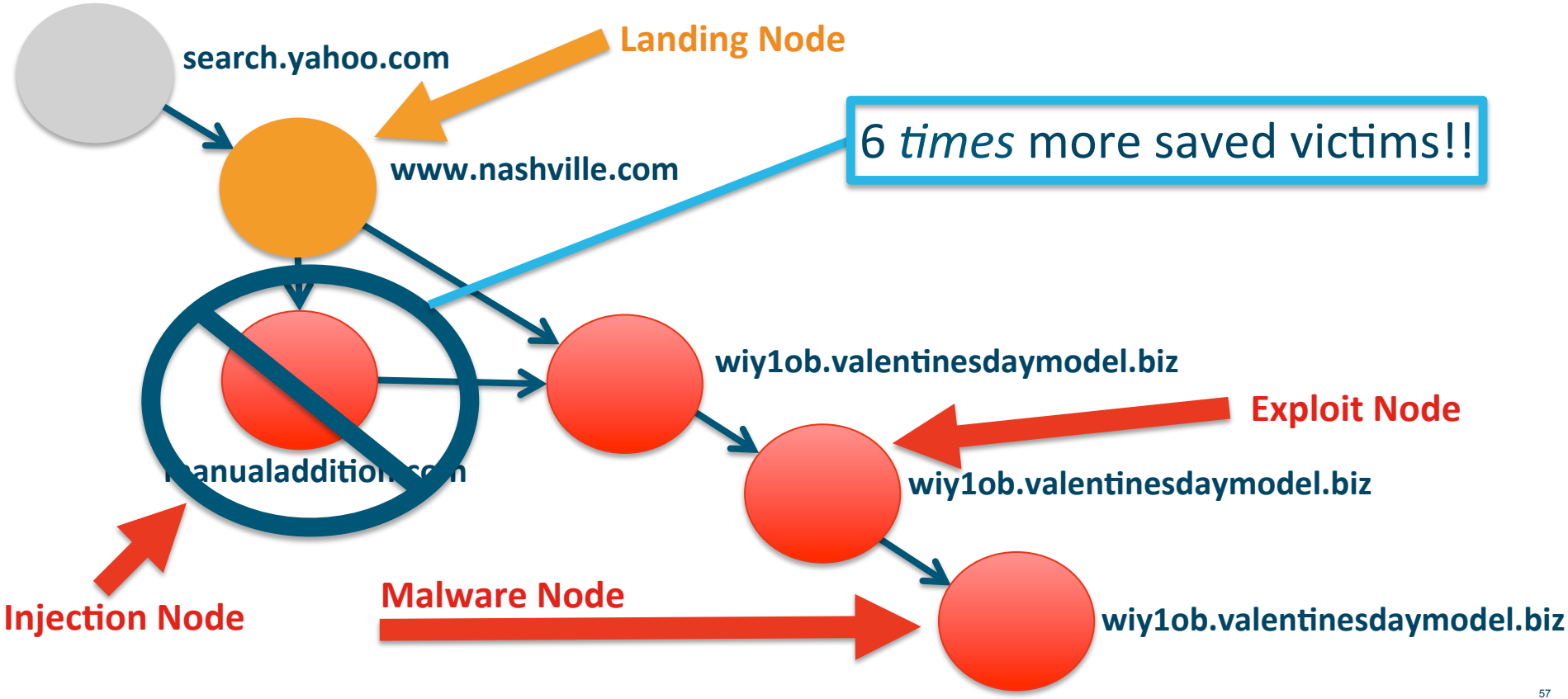
Single user's HTTP requests **5 minutes** prior to malware download.



Annotated Drive-by Download Path



Drive-by Download Defense



Questions?

WebWitness: Investigating, Categorizing, and Mitigating Malware Download Paths

Terry Nelms[1,2], Roberto Perdici[3,2], Manos Antonakakis[2], Mustaque Ahamd[2,4]

[1]Damballa, Inc.
[2]Georgia Institute of Technology
[3]University of Georgia
[4]New York University Abu Dhabi

tnelms@damballa.com, perdisci@cs.uga.edu, manos@gatech.edu, mustaq@cc.gatech.edu