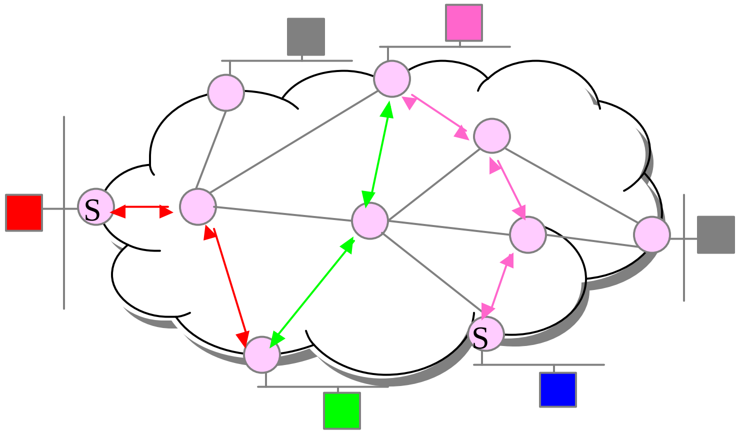
Tracing network attackers by encrypted stream matching

Florian Buchholz, Joshua Jenks, Jamie Van Randwyk
PI: Clay Shields
Purdue University, CERIAS

Problem

- Internet attackers often establish a chain of connections from one compromised host to another across the Internet
- This technique is used by attackers in an attempt to hide the source host from which they actually logged in, to reduce the chances of being caught

Connection Chain Diagram



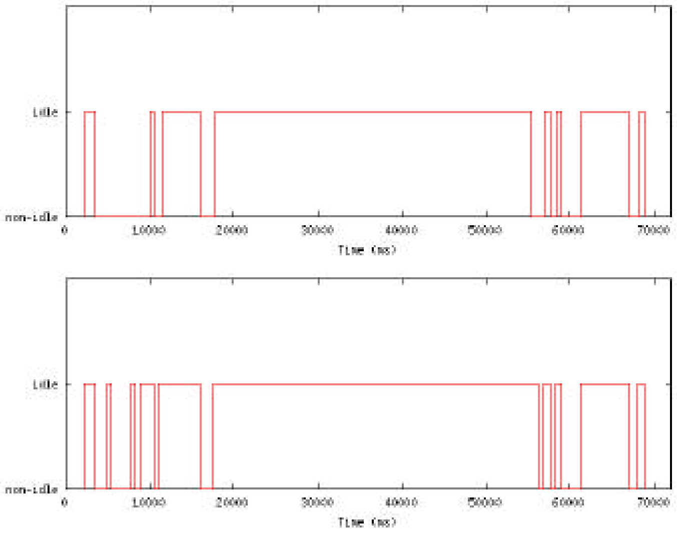
Goal

- Match a TCP stream to other streams in a global Internet to determine whether or not they are part of the same connection chain
- Known as **stream correlation**
- Eventually trace an “attack” stream back to its origin

Our Approach

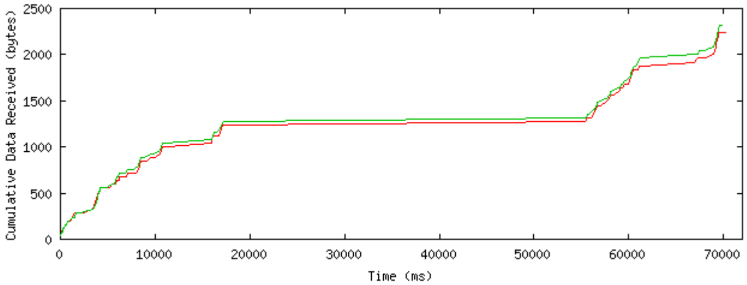
- Record data streams at various hosts
- Generate a thumbprint of each stream using:
  - Idle times between arriving datagrams (attacker’s “think time”)
  - Number of bytes received per time period

Idle Times



Idle time comparison between source (top) and destination (bottom)

Traffic Sizes



Cumulative traffic sizes seen by source and destination

Conclusion

- Correlating idle times appears more promising than traffic size
- Matching techniques being used need refining
- Defeat the idle time correlation technique by injecting traffic during idle times using shell scripts or modified ssh client (for which code has been written)