Mining and Detecting Connection-Chains in Network Traffic

Ahmad Almulhem and Issa Traore

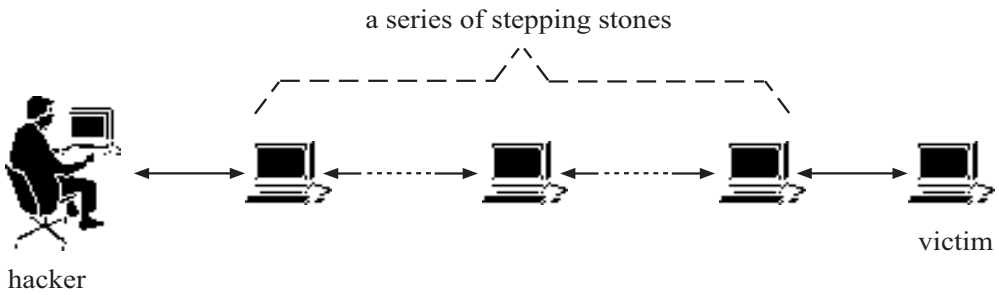
Department of Electrical and Computer Engineering
University of Victoria

IFIPTM'07 – Moncton, Canada
July 31, 2007

Outline

Outline

Connection Chains



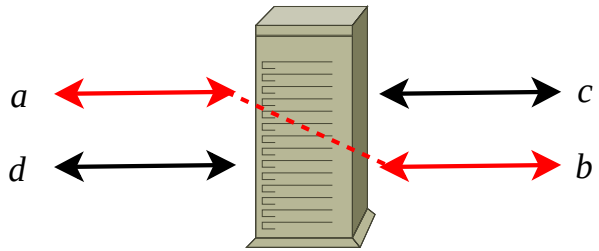
Connection chain

A set of connections created by sequentially logging into a series of hosts, known as stepping-stones [Staniford-Chen 1995, Zhang 2000].

- Used to carry attacks indirectly.
- Avoid disclosing true origin of attack.
- Stay anonymous!

Outline

Related Work



- Two types of host-based approaches proposed:
  - search processes [Carrier and Shields (2004)], [Kang et.al. (2004)].
  - modify OS [Buchholz and Shields(2002)].
- Disadvantage:
  - OS specific.
  - searching processes may fail.
  - modifying OS is costly and may break software.

Outline

Association Rule Mining

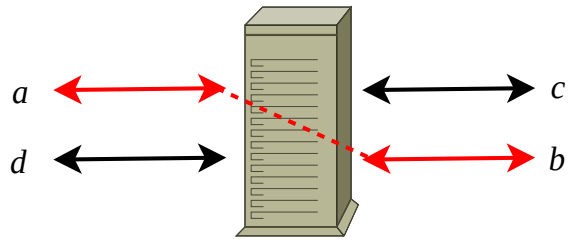
Definition

Association Rule Mining refers to a methodology that is is used to discover interesting relationships in large data sets [Tan 2006].

- $I = \{i_1, i_2, \ldots, i_n\}$ is a set of items.
- $T = \{t_1, t_2, \ldots, t_N\}$ is a set of transactions, where $t_i \subseteq I$.
- An itemset X is defined as a set of items; $X \subseteq I$.
- An association rule is an implication of the form $X \rightarrow Y$, such that $X \bigcap Y = \phi$.
- The strength is measured by:
  - Support s: X and Y occur together in $s\%$ of the total transactions.
  - Confidence c: Of all the transactions containing X, $c\%$ also contain Y.

Outline

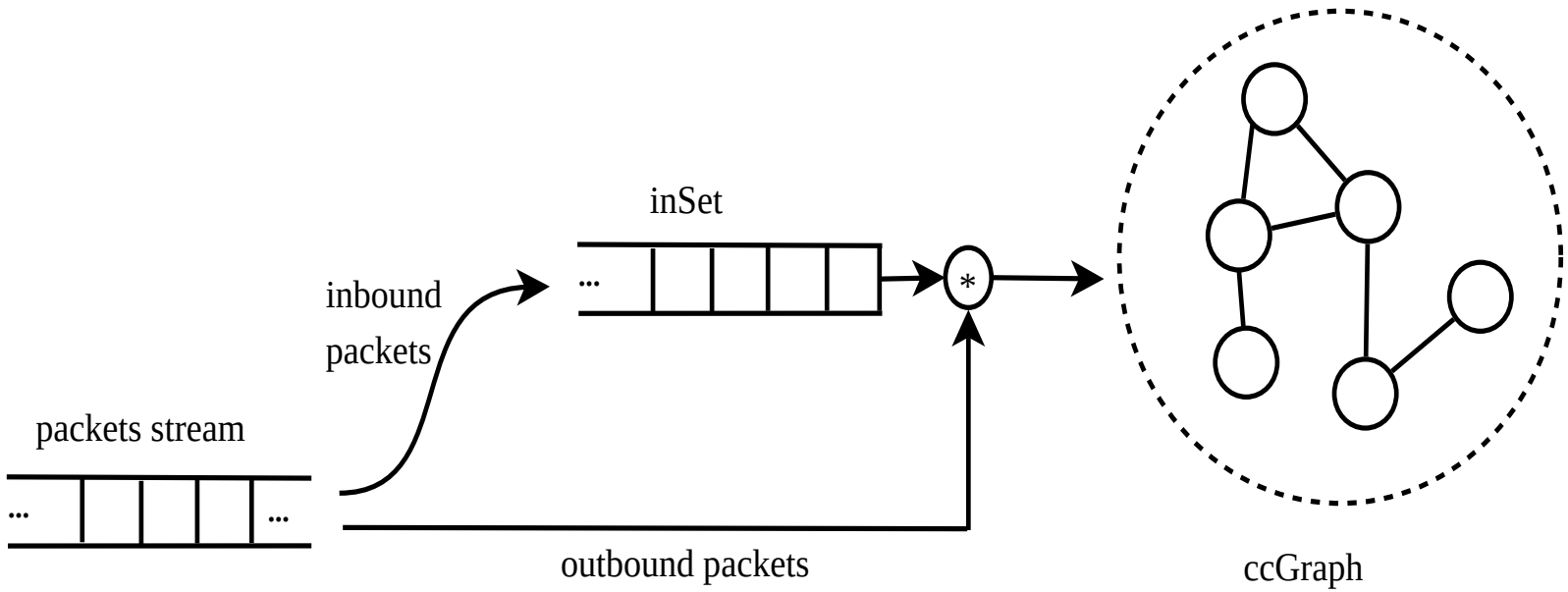
Mining for connection chains



Δ

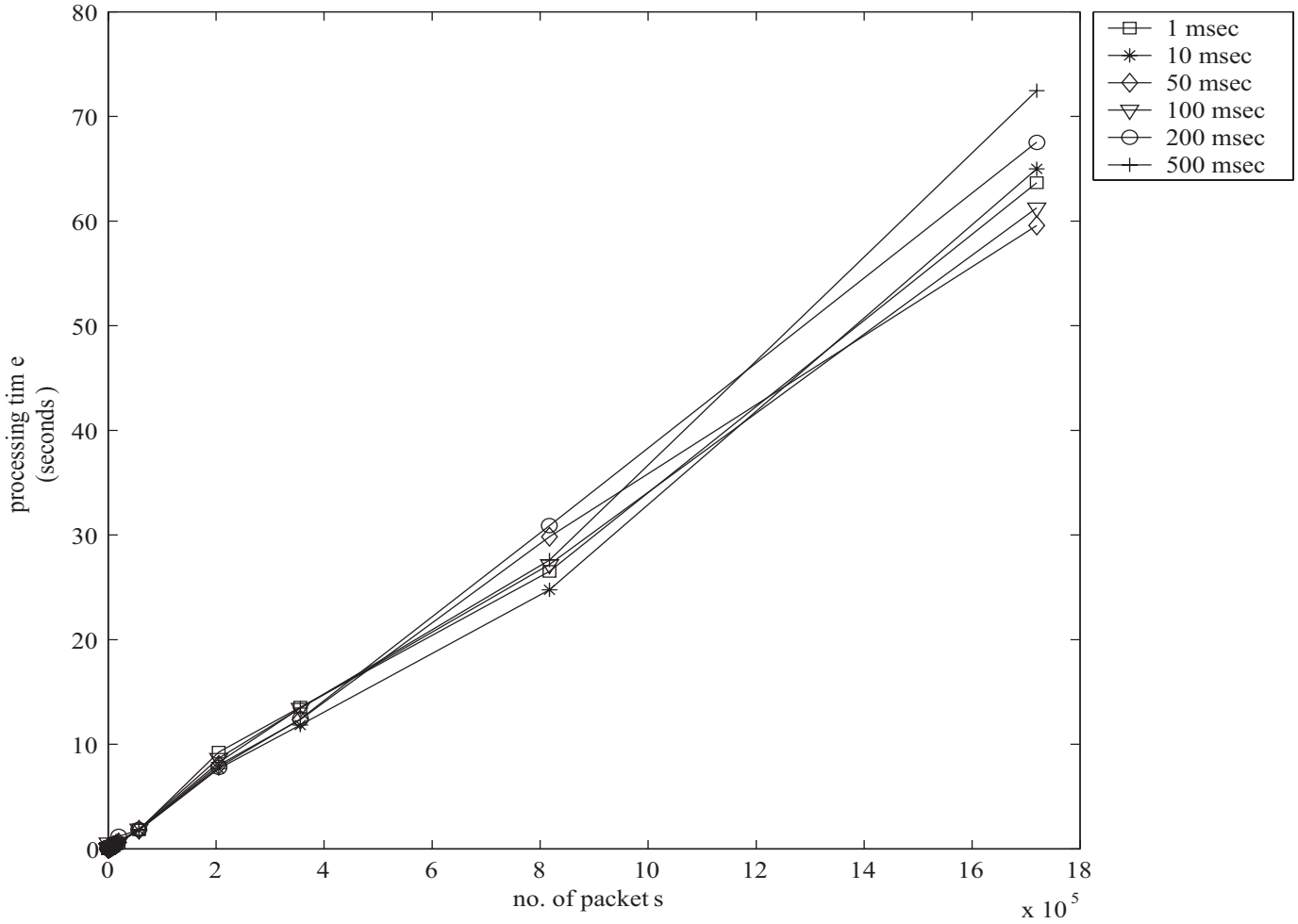
- items are connections.
- association rules are connection chains.
- $C = \{a, b, \ldots\}$; a set of connections.
- For two connections a and b, dynamically generate transactions:
  - input transaction; i.e. $[a]$ or $[b]$
  - chain transaction; i.e. $[a, b]$
- confidence$(\{a, b\}) = \frac{\sigma([a,b])}{\sigma([a])+\sigma([b])}$
  where $\sigma()$ is the support; how many times a transaction occurred.

Algorithm


inSet
inbound
packets
packets stream
*
outbound packets
ccGraph

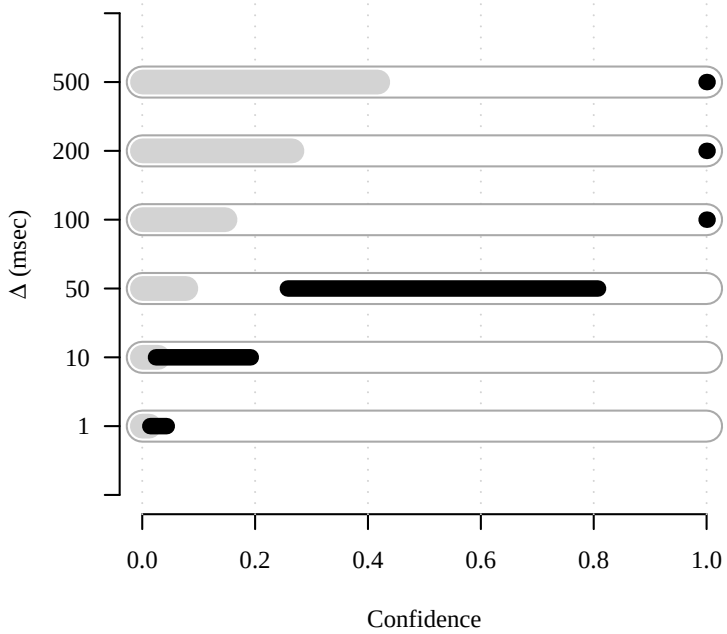
Outline

Evaluation (Processing Time)



Evaluation (Detection)

- Original trace has 1.7 million packets and 236 remote addresses.
- 88 simulated connection chains.
- Total: $236 + 88 = 324$; $\binom{324}{2} = 52326$ possible connection chains. Only 88 are true ones ($\approx 0.2\%$).



Outline

Summary

- A connection chain refers to the set of connections created by sequentially logging into a series of hosts.
- Attackers use them to stay anonymous.
- Proposed approach applies association rule mining.
- A confidence measure for the strength of a connection chain.
- Processing time is suitable for real-time.
- Efficient detection.