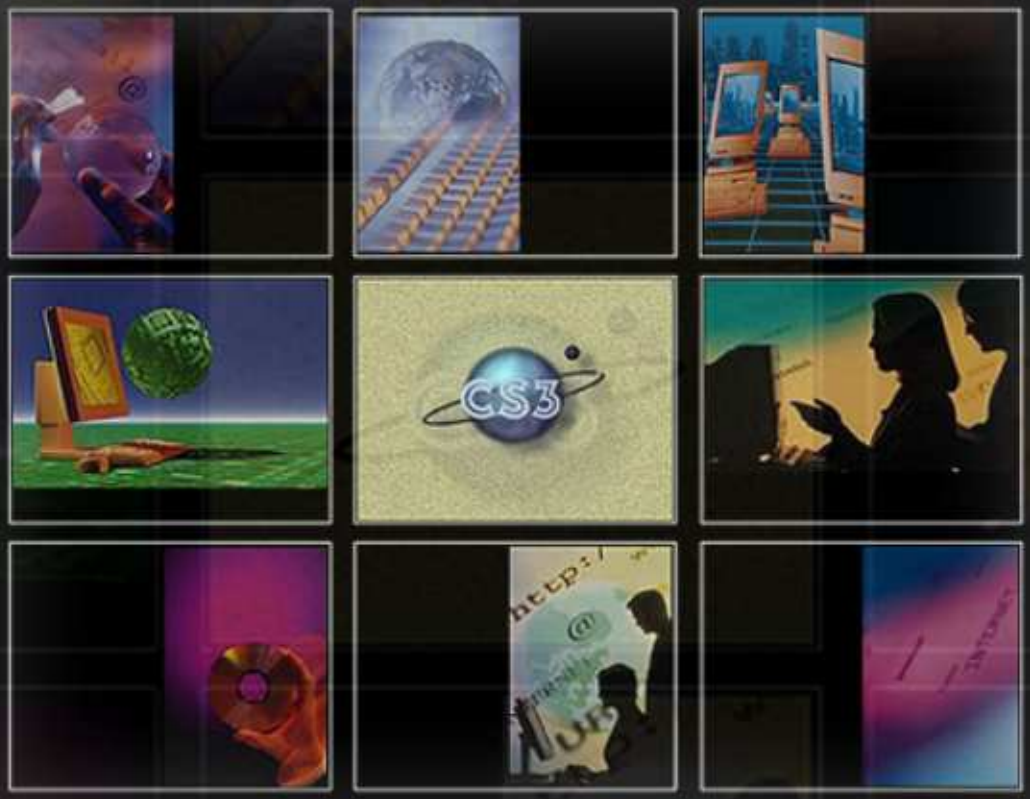
StarDeck Overview



Cs3, Inc.

http://www.cs3-inc.com

STARDECK

An Approach to the Attribution Problem

The problem at two levels.

Level 1:

Given a possibly spoofed, single IP packet, determine the possible IP addresses of the machines that could have generated the packet.

Level 2:

Determine if the actions of machine of origin (Level 1 attribution), are being caused by or controlled by activity at other machine(s) and identify such machine(s).

STARDECK : An Attribution Solution

TECHNICAL APPROACH

- Deploy remote traffic sensors to gather advance reconnaissance.
- Develop efficient traffic summaries to store sensor data indefinitely.
- Gather "*Reverse Routing*" data for the network to compute the possible origins for traffic at each remote sensor link.
- Build a table of origins for different sensor link identification signatures.
- Incorporate network intrusion data and application logs from any source as available.

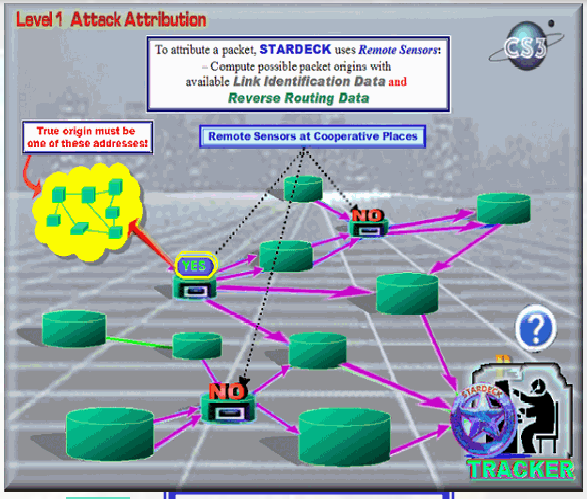
- **STARDECK Level 1 Attribution scheme for packet P:**
  - **Find the link identification signature for P from summaries of sensor data;**
  - **Look up the table of origins for that link identification signature.**
- **STARDECK Level 2 Attribution:**
  - **Incorporate new and existing heuristics for stepping stone control using traffic summaries;**
  - **Provide general query and correlation facilities for attribution questions.**

STARDECK: Systematic Tracking of Attackers using Routing Data and Event Correlation Knowledge

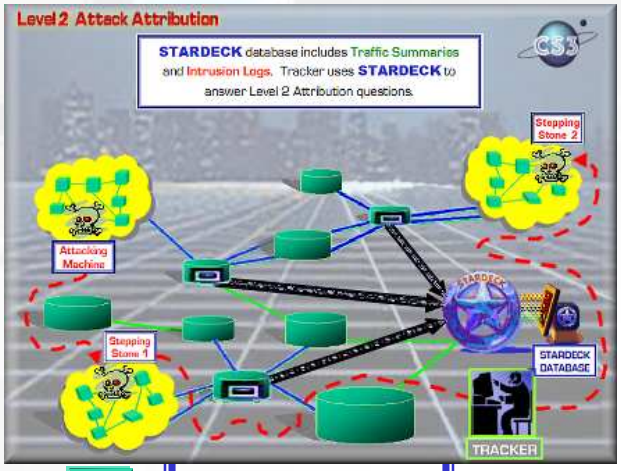
STARDECK : An Attribution Solution

INNOVATIONS

- Method for Level 1 attribution works with sparse cooperation – prior methods require universal adoption.
- Technique for Level 1 attribution can combine evidence from any of the prior methods wherever they are deployed.
- Efficient traffic summarization method that enables advance reconnaissance for arbitrary periods of time.
- Reflection Probe Method (RPM) to gather reverse routing data for large networks and for the Internet.
- Ability to answer Level 2 attribution queries beyond stepping stones to include zombies and worms.



L1 Animation



L2 Animation

STARDECK : An Attribution Solution

RESULTS

- Scalability of attribution solution shown analytically.
- Deployable solution (CONOPS developed).
- Robust with respect to many potential countermeasures.
- Patent application in consideration
- Payload independence has positive policy implications.

- **Prototype demonstrated "proof of concept":**
  - ✓ Deployment of traffic summaries from remote sensor data;
  - ✓ Ability of Reflection Probe Method to find reverse routing data;
  - ✓ Integration and correlation of network intrusion data;
  - ✓ Ability of tracker to find packet of interest using summaries;
  - ✓ Level 1 Attribution from a single packet under sparse cooperation;
  - ✓ Level 2 Attribution by answering queries related to machine control.

Options for Remote Sensors

- **Cooperative machines can observe traffic and communicate data to the tracker:**
  - Cost trade-offs: processing, storage, and communication –
    - **Send raw data:**
      - Least local cost, maximum communication cost
    - **Send summarized data:**
      - Time and space to summarize locally, less data to send
    - **Have databases at each sensor:**
      - Maximum time and space locally, least communication cost – only send answers to tracker questions
- **Cooperative routers can alter forwarded packets:**
  - e.g., encapsulating in tunnels to indicate packet origin
  - Adds processing cost inside router and bandwidth – specifics will depend on the scheme used

Getting Reverse Routing Data

- **New "Reflection Probe Method" (*RPM*):**
  - Goal: Find the remote sensor "signature" of packets sent from **X** to **Y**
    - Method: Sends packet from **Y** to **X** requesting reply
    - Record the sensors that observe the reply packet
  - Requires minimal cooperation (*reflection*) from **X**
- **Various other methods that require more cooperation:**
  - *Cooperating routers* can provide their routing tables
  - *Autonomous Systems* provide their BGP data to **STARDECK**
  - Places using *ingress filters* provide filter descriptions
  - Cooperating places use *traceroute* or remote sensors to sense the routes from themselves to the tracker
- **Partial data from above methods can be combined.**

Traffic Summaries

- **Goal: EFFICIENT traffic summaries for attribution**
  - Seeking compression factor of 100 or more:
    - Adequate to reduce transmission cost
    - Achieves acceptable storage cost
  - Minimal cost and effort to obtain the data.
- **STARDECK's Traffic Summarization Approach:**
  - Group traffic into "flows" (similar to NetFlow)
  - Includes more attributes specifically useful for attribution:
    - Timing data on periods where no traffic is observed
    - Protocol-specific – (for TCP, alternating payload volume)
    - Header field values appear & frequency (TTL, TCP flags…)
  - Multiple flows are grouped together for better compression
  - Unlike NetFlow, summarizes "abnormal" (attack) data.
- **STARDECK can use NetFlow data where possible:**
  - Not as good for attribution, but more widely available.

Level 2 Attribution

- **STARDECK Database incorporates:**
  - Traffic summaries from remote sensors including NetFlow data where available; and
  - Intrusion or application log data as available.
- **Specific Level 2 Correlation Heuristics:**
  - Traffic summaries directly show some stepping stone control:
    - e.g., Connection to Port 22 indicates *ssh* control;
    - Netflow summaries are adequate for this case.
  - Correlations based on timing and quantity of data transmitted (*quiet-time, tcp-turns, etc.*) also can point to stepping stone control
    - Netflow is *not* adequate to support this heuristic.
  - Specific attacks/modes of control recorded in intrusion logs can be correlated with traffic summaries to characterize traffic used in such attacks/control modes elsewhere.
- **Tracker can formulate general queries related to machine control and attribution as needed.**

STARDECK: Concept of Operations

PREPARATORY STEP IS TO GET SOME COOPERATION:

- Identify places to deploy advance sensors for surveillance
- Add new sensors as more cooperation becomes available

ONGOING GATHERING OF DATA by STARDECK:

- Traffic data gathered from available remote sensors
- Traffic data summarized, summaries stored in the database
- Reverse routing data gathered & stored in database
- Intrusion & application log data imported into database

WHEN THE HUMAN TRACKER IS INVESTIGATING AN INCIDENT:

PERFORM LEVEL 1 ATTRIBUTION:

Step 1: Use traffic summaries to find the packets of interest using observed properties of the attack.

Step 2: Use STARDECK to get the Level 1 Attribution result:
Quality of result will improve with additional cooperation.

PERFORM LEVEL 2 ATTRIBUTION:

Step 1: Use STARDECK heuristics to see if evidence of machine control was found in the database.

Step 2: Formulate general database queries to verify any specific theory or suspicion of whether machine control occurred.

Validation of STARDECK

- **Analysis of Scalability Issues:**
  - Traffic Summaries:
    - Compression factor of 100 adequate:
      - Back of envelope estimate: summary of all Internet traffic would take 1000 100GByte disks per day.
    - Current Summaries can typically meet the above goal;
    - Summarization of high-speed traffic feasible, but may not be cheap.
  - Getting Reverse Routing Data for the Internet –The BGP Hypothesis:
    - Autonomous System (AS) paths *from* each address in a BGP block to a given destination are identical;
    - Reverse routing table for entire Internet is 100K x 100K;
    - RPM can gather Internet-wide reverse routing data.
- **Metrics for Attribution Result:**
  - Inaccuracies are introduced into Level 1 attribution:
    - Instability of Routing Data;
    - Measuring the ongoing validity of the BGP hypothesis.
  - Inaccuracy metrics being developed in this project;
  - Attribution result will be presented with the metrics for inaccuracy to provide the tracker perspective.

STARDECK Countermeasures

- **Level 1 Attribution:**
  - Attacker controls cooperating remote sensors:
    - Can provide false data to mislead STARDECK.
  - Attacker controls routers:
    - Reverse routing data gathered via RPM assumes that attack packets route the same way as reflection probe packets.
  - Changes in routing can lead to incorrect results if they are not detected
    - STARDECK maintains routing data at periodic intervals.
  - STARDECK designed to rely on data that is harder for attackers to control:
    - Will work even if packets are spoofed or if hosts are controlled by attacker.
- **Level 2 Attribution:**
  - Anonymization makes it much more complex.
- **General Barriers to Attribution:**
  - Onion routing – might present data gathering problems at the remote sensors;
  - NAT – At best, one can attribute up to the device doing NAT, but not beyond that.