

Attribution and Aggregation of Network Flows for Security Analysis

Annarita Giani
Ian De Souza
Vincent Berk
George Cybenko

Institute for Security Technology Studies
Thayer School of Engineering
Dartmouth College
Hanover, NH

FloCon 2006, Portland, OR

Why flow data

The context in which we are interested in flow analysis is the following.

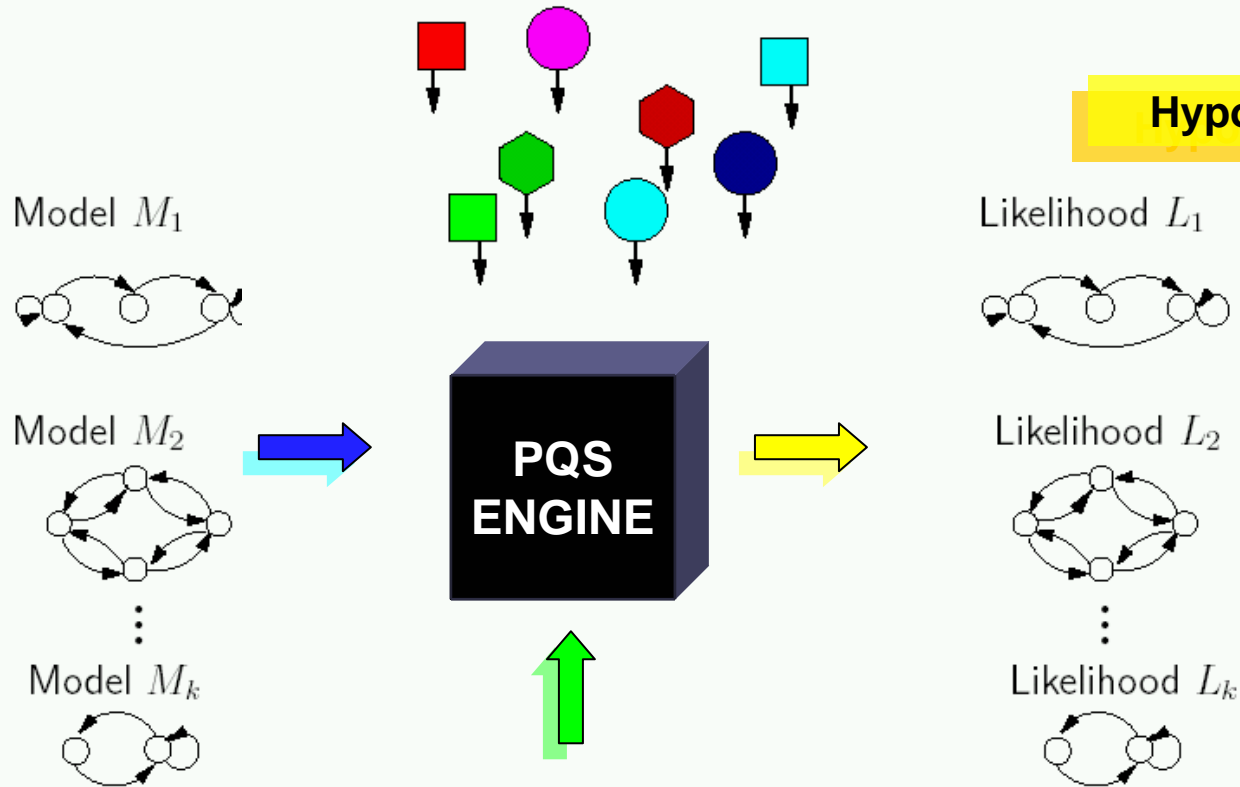
- We believe that **automated correlation** is hard to do.
- The world consists of **processes** so our approach to correlation is process-based..
- Introduction, in 2003, of generic process-based correlation engine concept and implementation, **Process Query System (PQS)**.
 - Integration of multiple existing and new sensor types and attacks models
 - Flow aggregation and correlations between flow data with security events
 - Implementation of a **PQS based process detection for Cyber Situational Awareness.**
 - **Need for flow data.**

Process Query System

Observable events coming from sensors

Models

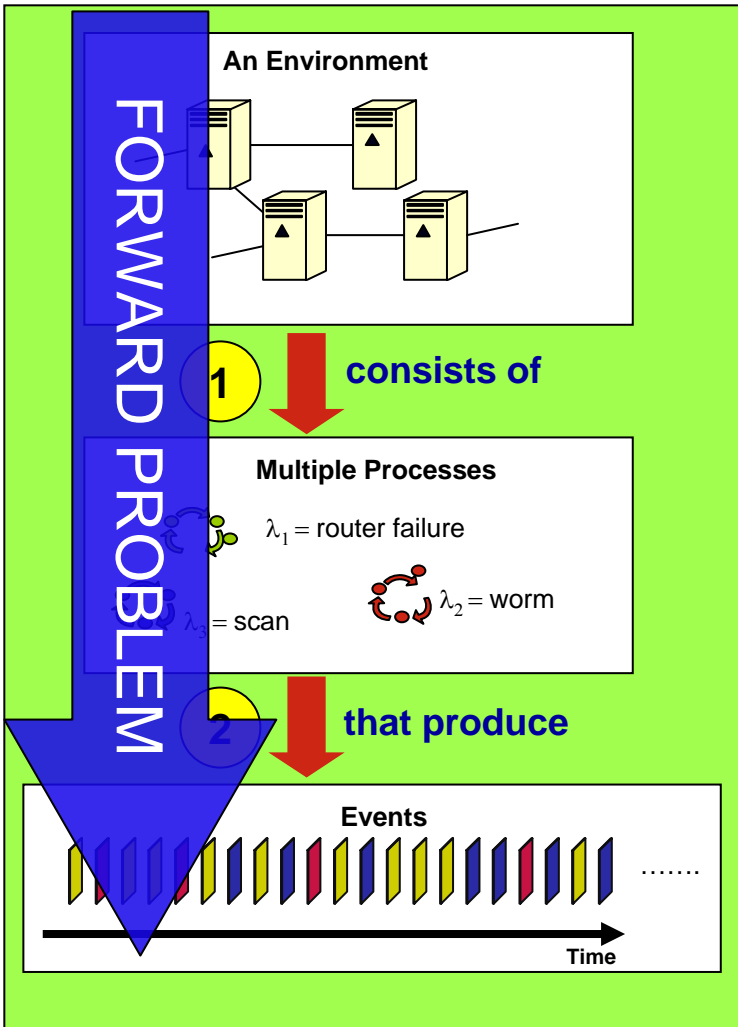
Hypothesis



Implemented for:
Vehicle Tracking
Computer Security
Social Network
Plume Tracking

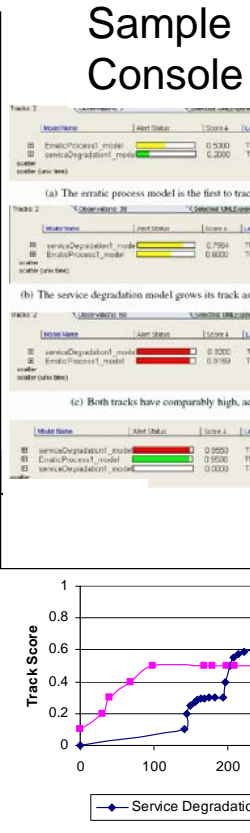
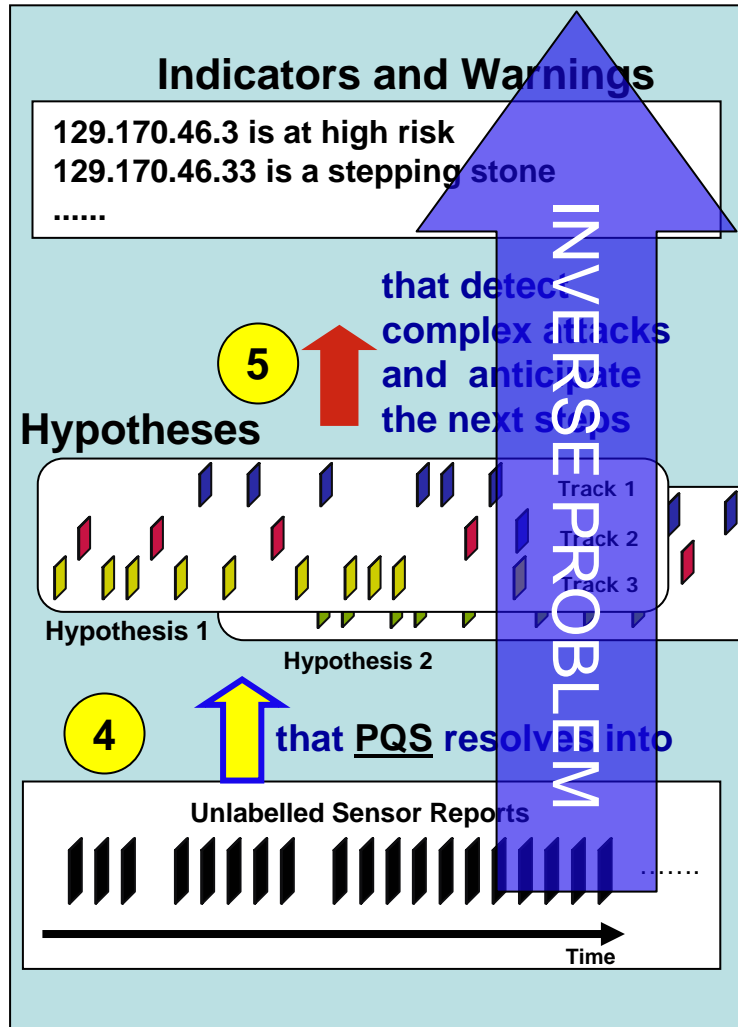
TRACKING
ALGORITHMS
Tracking Algorithms

Cyber Situational Awareness

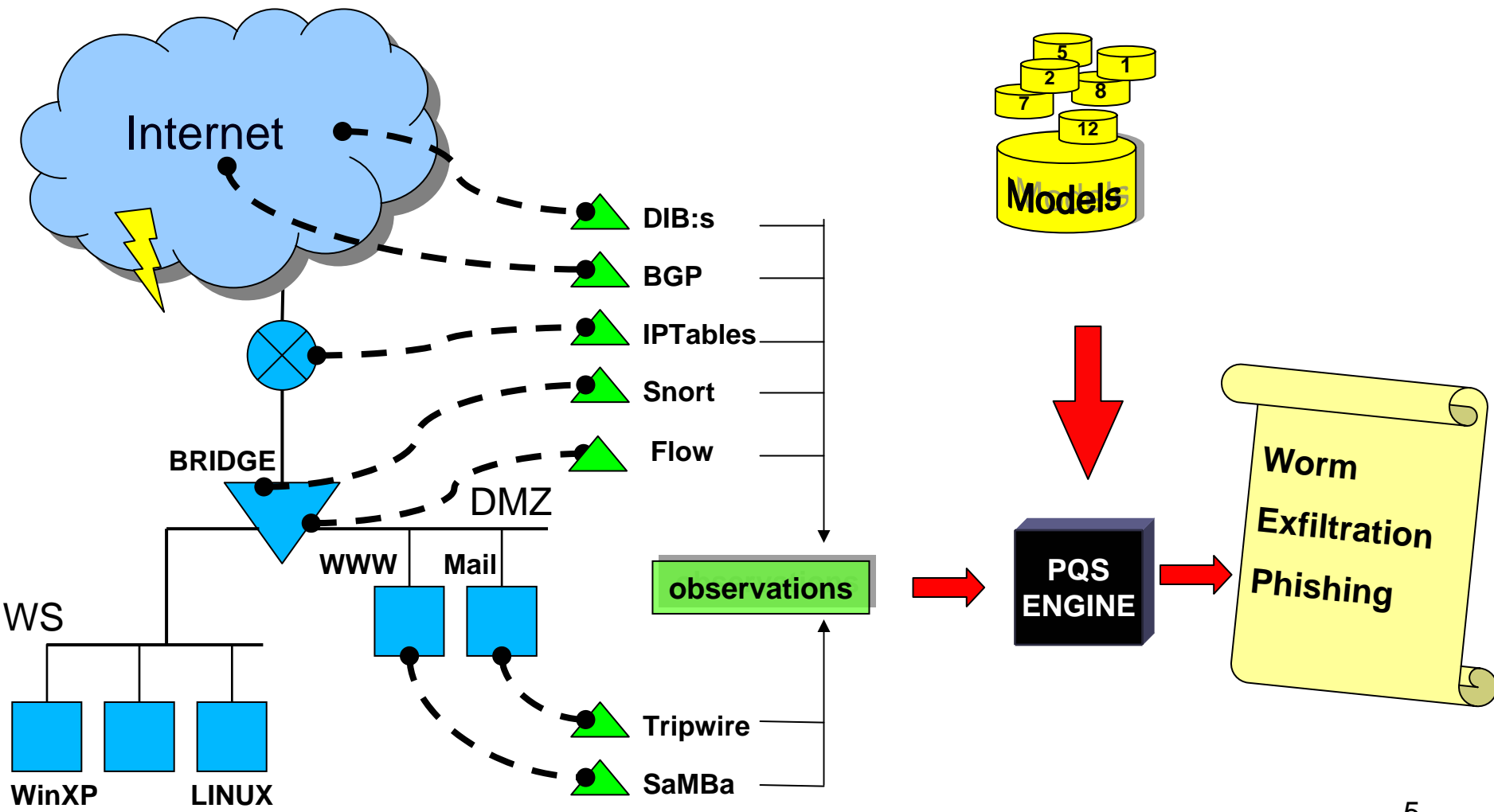


6 that are used for control

3 that are seen as










PQS in Computer Security










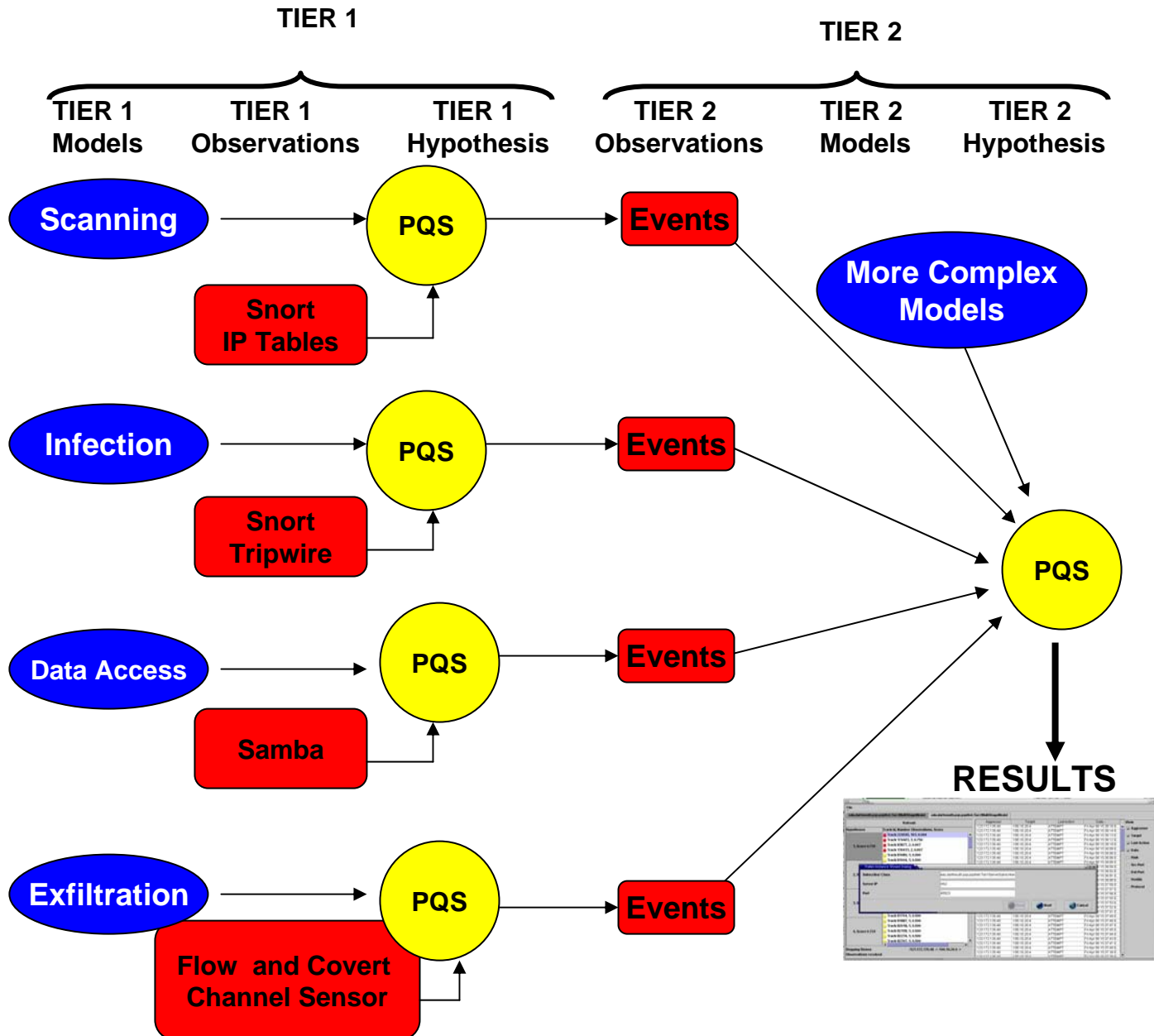
Sensors and Models

Sensors

-  **DIB:s** **Dartmouth ICMP-T3 Bcc: System**
-  **Snort, Dragon** **Signature Matching IDS**
-  **IPtables** **Linux Netfilter firewall, log based**
-  **Samba** **SMB server - file access reporting**
-  **Flow sensor** **Network analysis** ←
-  **ClamAV** **Virus scanner**
-  **Tripwire** **Host filesystem integrity checker**

Models

-  **Noisy Internet Worm Propagation – fast scanning**
-  **Email Virus Propagation – hosts aggressively send emails**
-  **Low&Slow Stealthy Scans – of our entire network**
-  **Unauthorized Insider Document Access – insider information theft**
-  **Multistage Attack – several penetrations, inside our network**
-  **DATA movement**
-  **TIER 2 models**

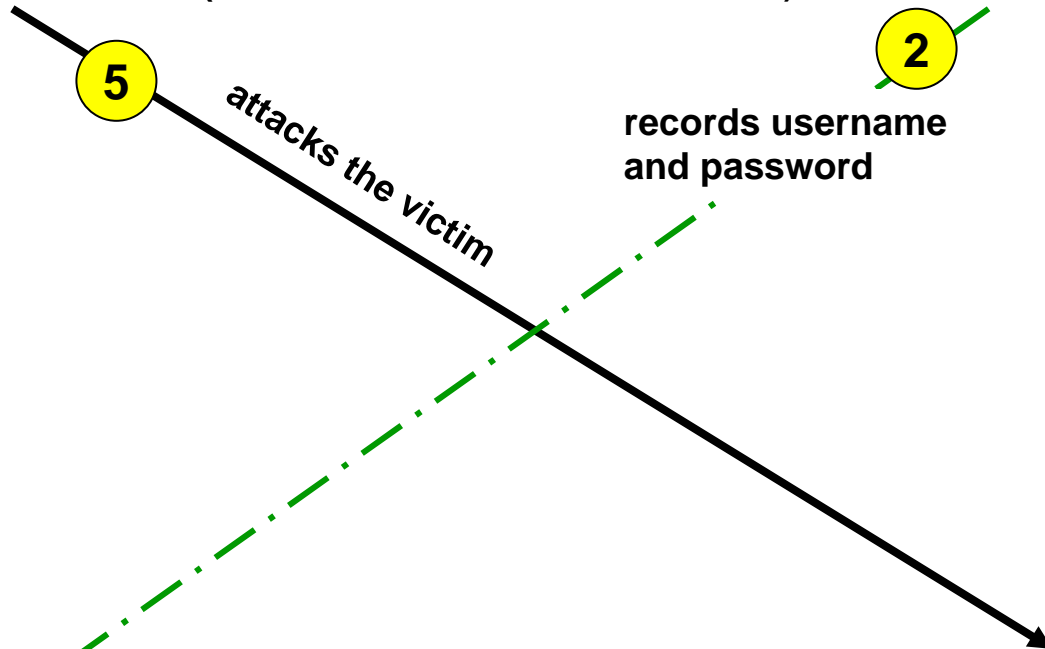
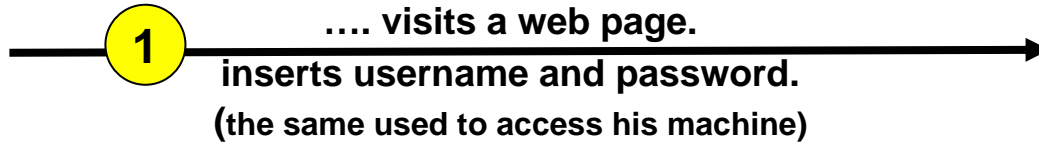


Multi Stage Attack Example: Phishing

Stepping
stone

... as usual browses the web and ...

Web page,
Madame X



100.20.3.127

165.17.8.126

accesses user machine using
username and password

uploads some code

Victim



100.10.20.9



3

4

6

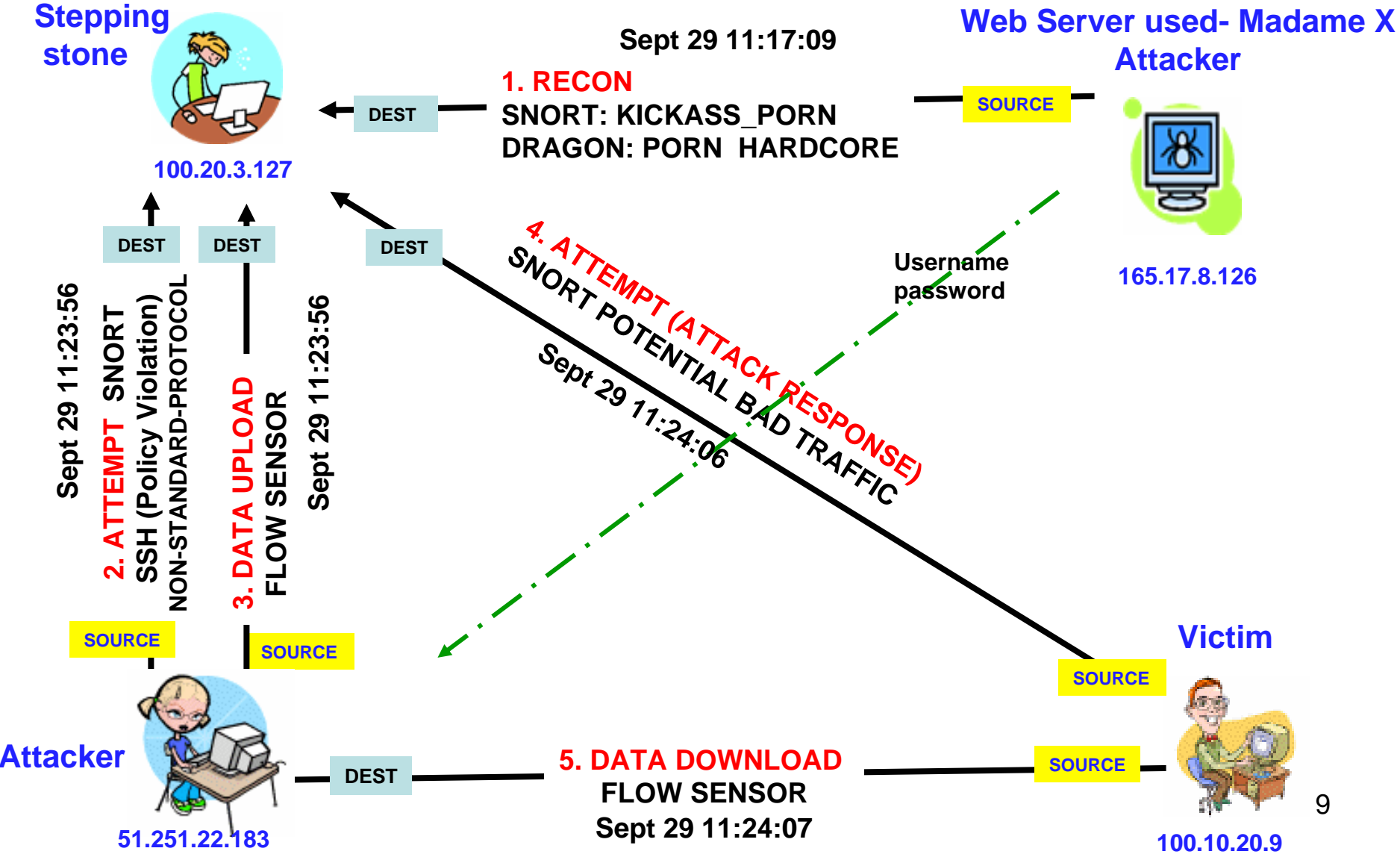
8

51.251.22.183

Attacker



Phishing Attack Observables



Flow Sensor

Based on the *libpcap* interface for packet capturing.

Packets with the same source IP, destination IP, source port, destination port, protocol are aggregated into the same flow.

- Timestamp of the last packet
- # packets from Source to Destination
- # packets from Destination to Source
- # bytes from Source to Destination
- # bytes from Destination to Source
- Array containing delays in microseconds between packets in the flow

Two Models Based on the Flow Sensor

Low and Slow UPLOAD

| Volume | Packets | Duration | Balance | Percentage |
|---|--|---|------------|---------------|
| Tiny: 1-128b Small: 128b-1Kb | 4:10-99 5: 100-999 6: > 1000 | 4: 1000-10000 s 5: 10000-100000 s 6: > 100000 s | Out | >80 |

UPLOAD

| Volume | Packets | Duration | Balance | Percentage |
|---|---|---|------------|---------------|
| Tiny: 1-128b Small: 128b-1Kb Medium: 1Kb-100Kb Large: > 100Kb | 1: one packet 2: two pckts 3: 3-9 4: 10-99 5: 100-999 6: > 1000 | 0: < 1 s 1: 1-10 s 2: 10-100 s 3: 100-1000 s 4: 1000-10000 s 5: 10000-100000 s 6: > 100000 s | Out | >80 |

Aggregation

Flow aggregation.

Recognizing that different flows, apparently totally unrelated, nevertheless belong to the same broader event (activity).

Flows are aggregated from captured network packets.

We aggregate flows into **activities**.

Example:

User requests a webpage (all DNS and HTTP flows aggregated)

Activity aggregation.

Recognizing that similar activities occur regularly at the same time, or dissimilar activities occur regularly in the same sequence.

We correlate activities into **activity groups, patterns**.

Examples:

- Nightly backups to all servers (each backup is an activity)
- User requests a sequence of webpages every morning.

Packet = Aggregated Bytes

Flow = Correlated Packets

Activity = Correlated Flows

Pattern = Correlated Activities

Web Surfing in Detail

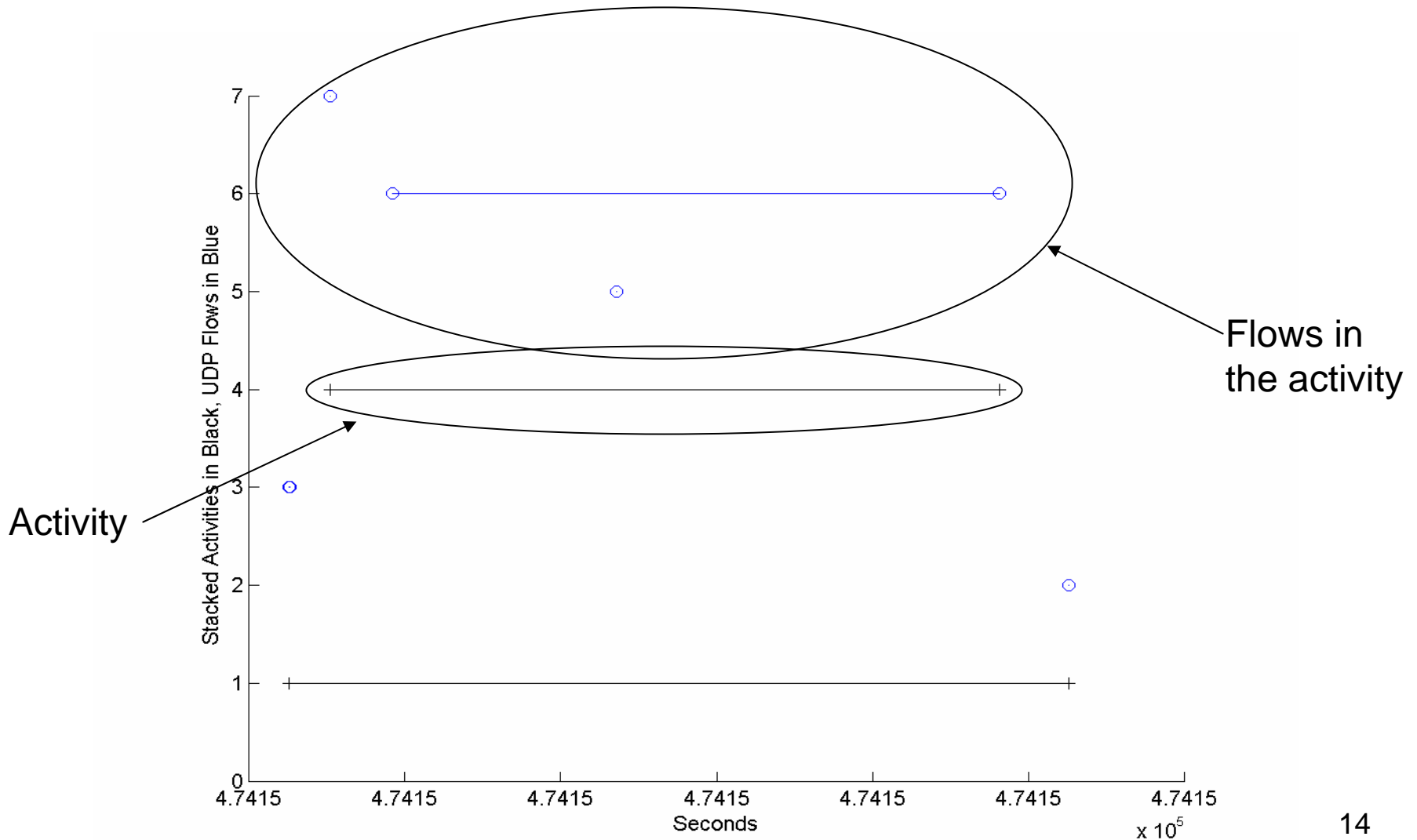
1. The **browser communicates with a name server** to translate the server name "www.dartmouth.edu" into an IP Address, which it uses to connect to the server machine.
2. The **browser forms a connection to the web server** at that IP address on port 80.
3. Following the HTTP protocol, the browser sends a GET request to the server, asking for the file "http://www.dartmouth.edu/index.html."
4. The web server sends the HTML text for the Web page to the browser.
5. The browser reads the HTML tags and formatted the page onto your screen.
6. Browser possibly initiates more **DNS requests for media** such as images and video.
7. Browser initiates more **HTTP and/or FTP requests for media**.

A FLOW IS INITIATED

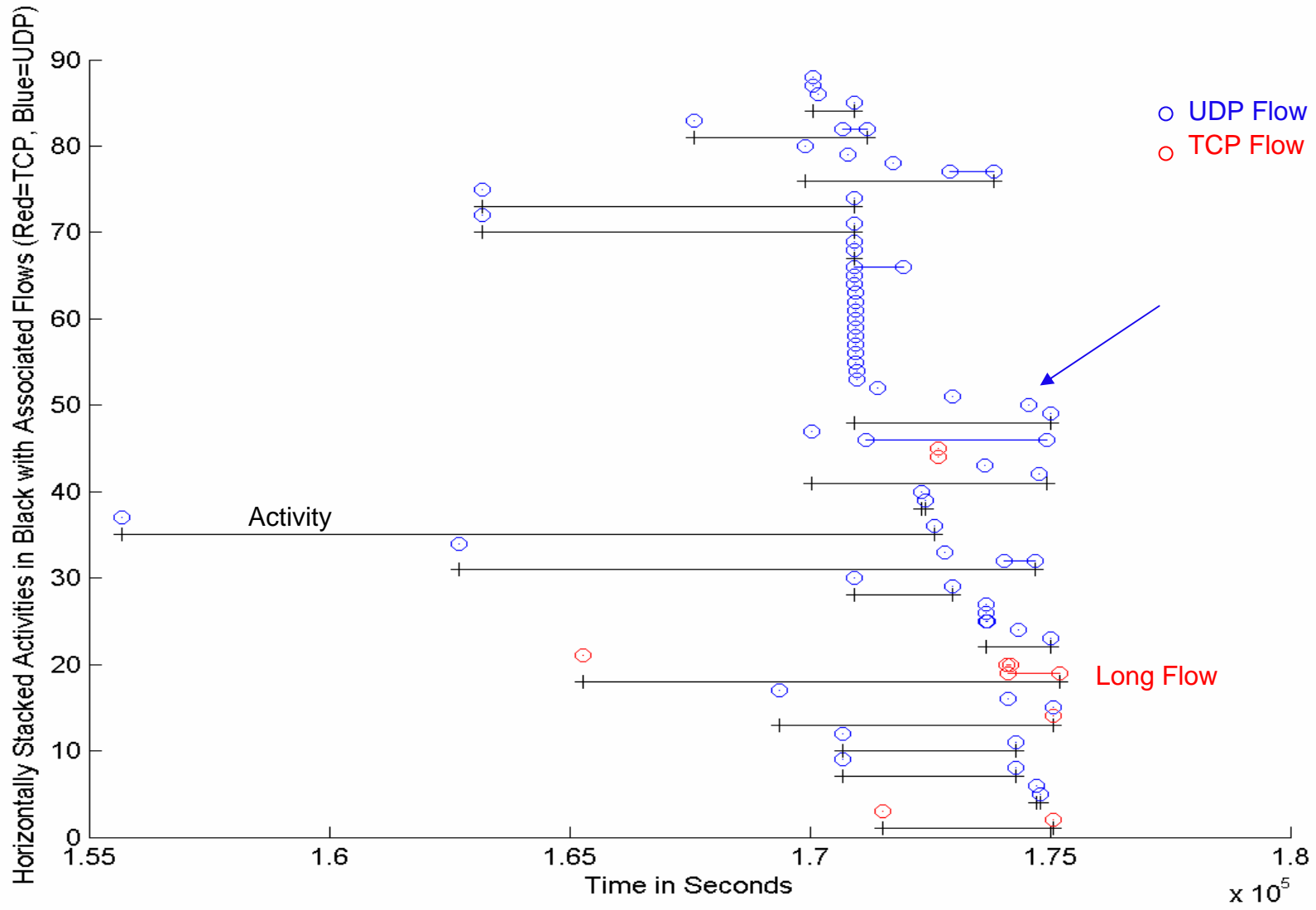
A FLOW IS INITIATED

MULTIPLE FLOWS ARE INITIATED...

Resulting Flows and Activity

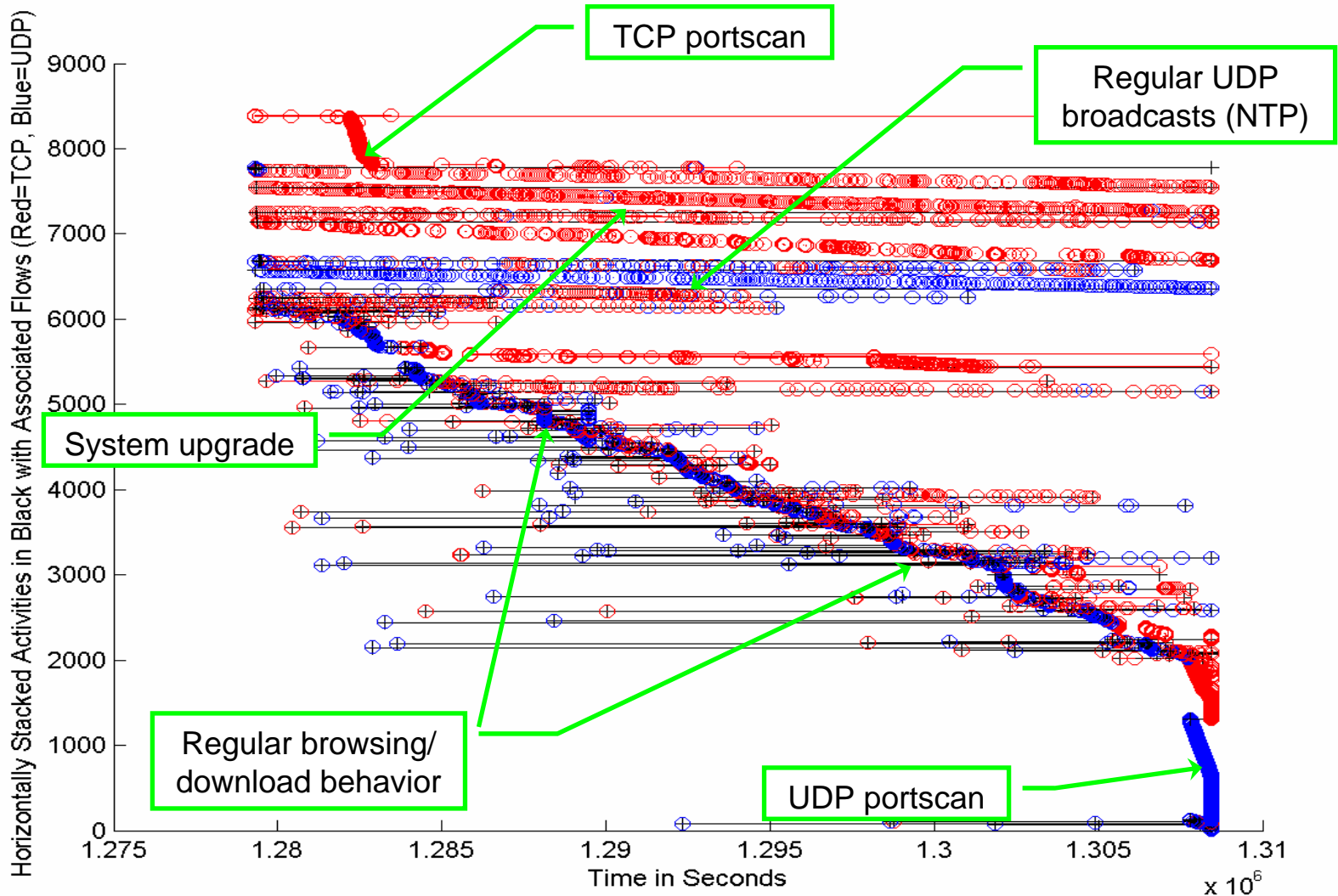


Activities and Flows



Complex Activities

Correlated
Network
Flows
Within
a LAN



Packets in a flow triggered IDS alerts

PQS instantiates models based on observation coming from flow and snort sensor.

Snort rule **1560** generates an alert when an attempt is made to exploit a known vulnerability in a web server or a web application.

Snort rule **1852** generates an alert when an attempt is made to access the 'robots.txt' file directly.

| Timestamp | Sensor | src IP | dst IP | Proto |
|-------------------|--------|---------------|-----------------|-------|
| Jul 09 16:28:32 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 16:29:35 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 16:44:44 | S1560 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 18:26:08 | S1560 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 21:05:03 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 22:31:08 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 09 22:31:08 | S1560 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 02:45:19 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 02:45:23 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 09:21:15 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 14:33:43 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 17:54:54 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 10 22:07:02 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 01:38:09 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 04:05:54 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 04:20:00 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 04:20:00 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 11:07:12 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 11:56:12 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| Jul 11 17:16:59 | S1852 | 65.54.188.140 | 208.253.154.195 | TCP |
| S Jul 10 02:30:27 | F | 65.54.188.140 | 208.253.154.195 | TCP |
| E Jul 10 23:55:56 | | | | |

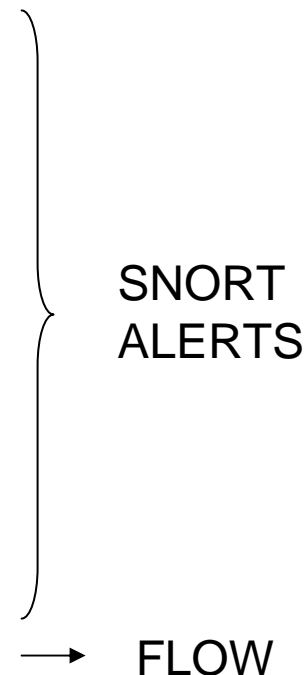


Table 2: A sample track of correlated IDS and Flow events

The flow can be characterized as malicious and further investigation must be done.

Future Direction

Theoretical approach for clustering aggregated flows.

Flow = As defined

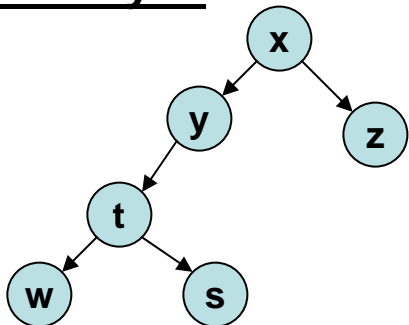
Activity = Aggregated flows

Pattern = Correlated Activities

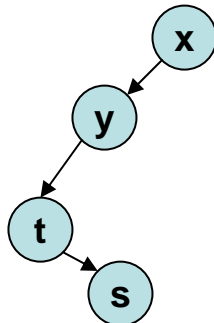
Approach: Graph theory (flows are the nodes and the edges are between correlated nodes).

We are thinking about defining a metric that captures the closeness between two different activities to allow grouping into patterns.

Activity 1.



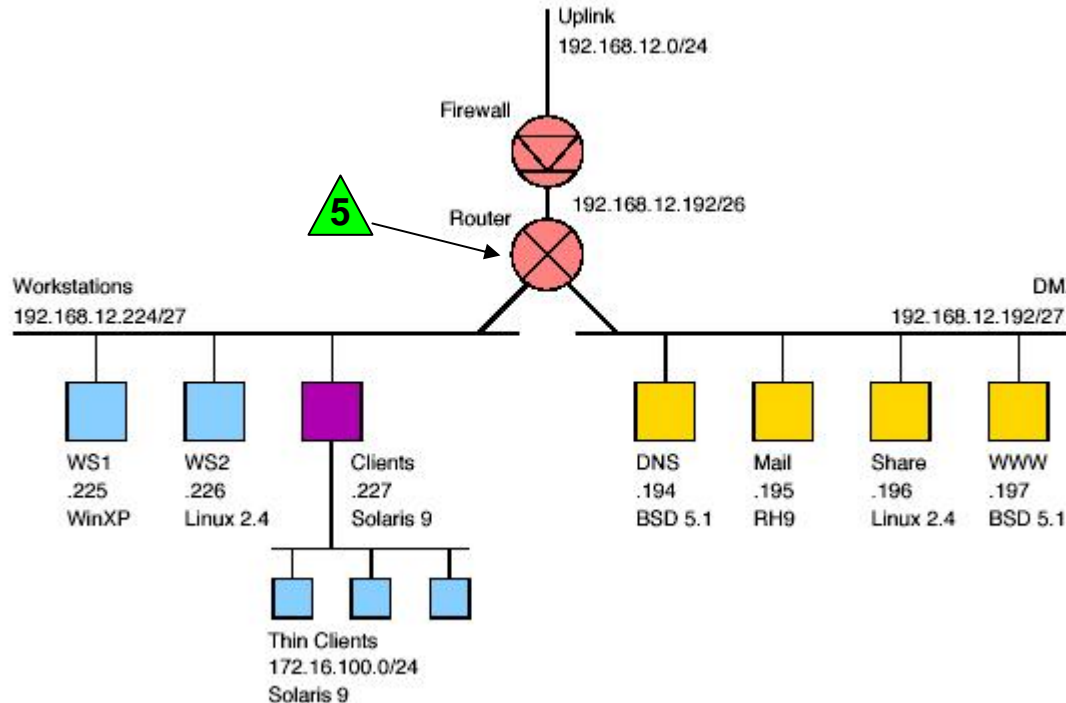
Activity 2.



Can they be grouped in one pattern?
Notion of distance between activities.

www.pqsnet.net
agiani@ists.dartmouth.edu

PQS-Net Network



Student and researcher use this network to browse the web, print documents, send upload and download files...

Web Surfing

208.253.154.210 host name
208.253.154.195 dns.pqsnet.net
129.170.16.4 ns.dartmouth.edu

1. ns.pqsnet.net requests www.nytimes.com ip address to ns.dartmouth.edu
2. ns.dartmouth.edu returns the ip address – 199.239.136.245
3. TCP three-way handshake between the host machine and the web server.
4. HTTP GET request to 199.239.136.245
5. TCP ACK from the web server
6. Other packets exchanges between the web server and the host

The image shows a Wireshark network traffic capture. The main pane displays a list of packets with columns for No., Time, Source, Destination, Protocol, and Info. A purple oval highlights a sequence of packets: a DNS query (No. 774), a DNS response (No. 775), an ARP request (No. 780), a TCP SYN (No. 782), a TCP SYN-ACK (No. 783), an HTTP GET (No. 784), and a TCP ACK (No. 785). Colored arrows point from the numbered list on the left to these specific packets. The packet details pane at the bottom shows the structure of the selected packets, including Ethernet II, Internet Protocol, User Datagram Protocol, and Domain Name System (query) fields.

All these network connections are related to the same host activity.