Network Monitoring for Performance and Security
The LOBSTER project

Evangelos Markatos
and Spiros Antonatos
Institute of Computer Science (ICS)
Foundation for Research and Technology – Hellas (FORTH)
Crete, Greece
Roadmap

- **Motivation**
  - What is the problem?
  - Why is it important?
- **Passive network traffic monitoring**
  - The SCAMPI and LOBSTER approach
- **Privacy protection**
- **How can you get involved?**
  - Follow our activities
  - Install a sensor
  - Become part of the Infrastructure
- **Demos/Applications**
Motivation

• Committee on Research Horizons in Networking (formed in 2001)
  – David Patterson, Chair, University of California at Berkeley
    • RISC processors, RAID storage, NOWs – clusters of workstations
  – David Clark, MIT Laboratory for Computer Science
    • The “father” of the “end-to-end” argument, on top of which the Internet design is based
  – Anna Karlin, Jim Kurose, Edward D. Lazowska, David Liddle, Derek McAuley, Vern Paxson, Stefan Savage, Ellen Zegura

• The committee was assigned to
  – “formulate a fresh look at networking research”

• They prepared a report
  – “Looking over the fence at networks: a Neighbor’s View of Networking Research”
  – They identified three “Grand Research Challenges”
The first GRAND Challenge in Computer Networking is to
“... develop and deploy the technology to make it possible to record a day in the life of the Internet...”

Committee on Research Horizons in Networking
Clark, Lazowska, Patterson, Paxson, Savage, Zegura, ....
2001

A day in the life of the Internet
The Challenge

• Why is it important?
  – “a data set with typical days for the next 10 years of the Internet might be a treasure chest for networking researchers”
  – Measurement-based GRAND challenges, such as the human genome, have served to
    • Crystallize research issues, and
    • Mobilize research efforts
  – Good data are necessary for the operational needs
    • Why is my network slow?
    • Which route do my packets follow?
    • Why is a particular flow missing lots of packets?
    • How much peer-to-peer traffic is there?

• Next GRAND Challenge in Networking Research:
  – Monitor a day in the life of the Internet
Since 2001, we are trying to make contributions towards facing this GRAND challenge:

- SCAMPI is an IST project: Funded by European Commission
  - Duration: 1/4/02-31/3/05

- LOBSTER is a Specific Support Action Funded by European Commission
  - Two-year project, Duration 1/10/05-31/12/06
What has been developed so far?

- A Network Monitoring Programming Environment
  - MAPI (Monitoring Application Programming Interface)
- An implementation of MAPI on top of several Monitoring cards
  - Regular network interfaces
    - Commodity Intel cards
  - Specialized network interfaces
    - Combo6 cards (by CESNET – Masaryk University)
    - DAG Cards (by Endace)
So, what is MAPI?

- Main Abstraction: The **Generalized NETWORK FLOW**
- A (generalized) Network Flow is a **subset** of the traffic:
  - All packets destined to **port 80**
  - All packets destined to [www.cnn.com](http://www.cnn.com) to port 80 containing the string “this is a test”
- Network flows efficiently define a **subset** of the traffic
- What can you do with the packets of the network flow?
  - Capture them – dump them to disk
    - Post-process them later
      - To find time distributions, to observe suspicious patterns
  - Apply functions to them, e.g.
    - Count them, count the number of bytes they contain, etc.
Example: Search for a worm

- **SLAPPER worm**: from port 2002 to port 80, contains string “|00 00|E|00 00|E|00 00|@|00|"

```c
int main() {
    // Create a network flow consisting of SLAPPER worm packets
    int fid;
    struct mapi_pkt *pkt;
    /* create a flow using the eth0 interface */
    fd = mapi_create_flow("eth0");
    /* the bpf part of the signature */
    mapi_apply_function (fd, "BPF_FILTER", "udp and src port 2002 and dst net 139.91.23 and dst port 80");
    /* the content search part of the signature */
    mapi_apply_function (fd, "STR_SEARCH", "|00 00|E|00 00|E|00 00|@|00|", 0, 100);
    /* must use TO_BUFFER in order to read full packet records */
    fid = mapi_apply_function (fd, "TO_BUFFER");
    /* connect to the flow */
    if( mapi_connect (fd) < 0) {
        printf("Could not connect to flow %d\n", fd);
        exit(EXIT_FAILURE);}
    while(1) { /* forever, wait for worm packets */
        pkt = mapi_get_next_pkt (fd, fid);
        printf("\n Slapper worm packet! \n");
        print_IP_pkt(pkt);
    }
}
```

A day in the life of the Internet       markatos@ics.forth.gr
I have more than one Internet connection
How can I monitor it?

• Install several passive monitors
• Would you like to see all of them as one?
  – Use DiMAPI (Distributed Version of MAPI)
• DiMAPI enables you to monitor several lines as if they were one:
• The SCOPE abstraction:
  – SCOPE is a set of lines (interfaces) to monitor
• `mapi_create_flow("host1:eth2, host2:/dev/dag0, host3:eth1");`
• Fully compatible with MAPI
  – All previous functions work as usual
My friends have also installed passive monitors
Can we collaborate?

• Yes!
• Distributed Passive Network Monitoring
• Issue queries to remote sensors
  – e.g. today I see lots of pings to ports 1515 and 1414
    • Do you also see them?
    – Let’s all gather our top-10 most popular ports to pinpoint any anomalies
• Issues queries to several of them at once
  – Using the “SCOPE” abstraction
• Correlate the results received
• Isolate and study performance problems
  • Troubleshooting
    – I see weird port scans from port 1414, and 1515. Is anyone else seeing them?
  • Gather statistics
    – How much Gnutella traffic do we have today?
  • Observed bandwidth for individual applications
    – How many Mbps does my GRID-enabled application receives today?
  • Identify covert channels
    – Is there any “covert” traffic masquerading as web traffic using port 80?
  • Identify BOTS
    – Are there any compromised computers in my network engaged in suspicious activities?
• Monitor applications with dynamic ports
  – Teleconferencing, p2p, etc.
What can I do with it?
Find out who generates all this traffic...

69% of the traffic is unaccounted-for
- Does not belong to static ports
- Maybe belongs to p2p applications that use dynamic ports
- Maybe belongs to media applications
- The bottom line is:
  - We don’t know
- LOBSTER can help us find the answer
  - Digs down into control packets
  - To find dynamic data ports
Potential LOBSTER applications: Early-warning systems

- Automatic detection and fingerprinting of zero-day worms
  - How?
    1. Find lots of “similar” incoming packets [ICC 05]
      - Which have not been seen before
    2. “Execute” the incoming request
      - If the incoming request executes for a long time it is suspicious [SEC 05]

- Use Passive Monitoring to Complement Honeypots
  - With Shadow honeypots: “stand-by servers” [USENIX SEC 05]
    • Network-level inspection: it finds “suspicious” packets
    • Shadow honeypots: receive “suspicious” packets and serve them
What do I need to start monitoring?

- A regular PC
  - With a regular network Interface
  - Put it in promiscuous mode
  - Mirror your traffic from a router to the PC
  - Start collecting packets

- What is the cost of the hardware?
  - Minimal – use an old PC
  - It works well for 10-100 Mbps lines
  - It even works for 1 Gbps lines for lightweight monitoring
    - not heavy string searching

- For higher speeds (10Gbps) you may need
  - A better PC (better I/O and memory bandwidth)
  - A better network Interface (Endace, Combo)

- But you can always start small, inexpensive, and grow into it
What software do I need?

- **Off-the-shelf libraries**
  - MAPI ([http://mapi.uninett.no/](http://mapi.uninett.no/))

- **Applications**
  - Some are provided with MAPI
    - Counting
    - DoS attack detection, etc.
  - Write your own
    - And share it!
How about my privacy?

- Concerned about the privacy of
  - My users, my network, my topology
- Before talking about privacy ask the following:
- Are you willing to share information?
  - **No, absolutely not. I will not reveal any information.**
    - I will not reveal even the load of my network
  - Maybe I can share some general statistics
    - Much like U of Wisconsin does
  - I am willing to share the headers of the packets (IP addresses anonymized, payload stripped)
    - Just like NLANR does today
  - I would like to share
    - all information with my local administrators
    - Some information with associated researchers
    - Anonymized information with the rest of the world
  - **Yes – share everything**

LOBSTER can support all of the above
Sharing - Anonymization

- Anonymization Framework
- Flexible of anonymization
  - Per-field anonymization
  - Examples
    - Anonymize src/dst IP addresses and strip the payload (NLANR traces)
    - Prefix preserving src/dst IP address anonymization and strip the payload (Dartmouth traces)
    - For FTP traces: strip ftp passwords, strip files names transferred (LBL ftp traces)
Anonymization

• **Flexible Anonymization**
  – Different users see different anonymization views
  – “local administrators” may be able to see more information than “outside” users

• **Authentication**
  – Users are authenticated using the Keynote mechanism
    • RFC 2704

• **LOBSTER** can help administrators express what they are willing to share and with whom
More Information on Anonymization

- Anonymization Framework Definition
How can you get involved?
Visit us from time to time

http://www.ist-lobster.org/

A day in the life of the Internet
How can you get involved?
Follow the LOBSTER activities

• Subscribe to LOBSTER news
  – http://www.ist-lobster.org/announcements/

A day in the life of the Internet
markatos@ics.forth.gr
How can you get involved?
Follow the LOBSTER activities

• Come to our workshop
  – Co-located with the TNC Conference in Catania May 2006

• 1st LOBSTER Workshop
  – Held on 7 June 2005 in Poznan, Poland
  – Speakers from FORTH, TNO, GEANT, CESNET, VU, ENDACE

A day in the life of the Internet
markatos@ics.forth.gr
How can you get involved?
Follow the LOBSTER activities

• Come to our passive monitoring tutorial
  – Held jointly with the TNC Conference in Catania, May 2006
How can you get involved?
Would you like to experiment with it?
Install a passive monitor

• How?
  1. Get a regular PC
  2. Read the FAQ
     • http://www.ist-lobster.org/about/faq.html
  3. Start capturing packets
• Receive help and feedback from LOBSTER
• Join the MAPI mailing list
  – http://mapi.uninett.no/
• Use it for your own monitoring purposes
How can you get involved?
Become a part of the LOBSTER team

• Would you like to share data?
  – Even highly anonymized data?
• Stay tuned
  – LOBSTER phase two is coming up:
  – In June 2006 the LOBSTER infrastructure will start accepting members
Collaborations

- **Symantec**
  - Worm detection

- **NRNs**
  - Node installation

- **GigaCampus**
  - >100 PM sensors in Norway

- **NLANR (USA)**

- **Far East**
  - Shanghai Jiao Tong University (Mao Weihua)
    - Interested in network monitoring collaboration
    - And common proposals
  - Institute for Infocomm Research (Singapore)

A day in the life of the Internet
Summary

• Networking GRAND Challenge:
  – “Monitor a day in the life of the Internet”
• LOBSTER/SCAMPI are about (Passive) Traffic Monitoring
• MAPI (Monitoring Application Programming Interface)
  – Write your own applications
• Concerned about privacy?
  – Answer the following question:
    • “How much data am I willing to share and with whom?”
  – LOBSTER can help you implement this answer
• How can you get involved
  – Just browse www.ist-lobster.org from time to time
  – Join our email list
  – Install a passive monitor for your network
  – Link your monitor to the LOBSTER infrastructure
Demo!!

- Worm hunting
- Accurate traffic categorization
- Covert traffic
Worm hunting

- Monitor the incoming network traffic to find
  - The Witty worm and the Blaster worm
Dynamic Port categorization

- Categorize FTP traffic
- FTP uses dynamic ports, so to find the data ports,
  - applications have to dig down in the payload of control packets
Covert Traffic

- Gnutella p2p over port 80
Back up slides
LOBSTER Project Objectives

- Deploy an advanced pilot Internet Traffic Monitoring Infrastructure across Europe.
- Realize the appropriate data anonymizing tools that will prohibit unauthorized tampering with the original traffic data.
- Develop novel applications enabled by the availability of the passive network traffic monitoring infrastructure:
  - Accurate traffic characterization for programs using dynamic ports.
  - Spread of zero-day worms.
  - European Internet measurement service.
- Provide anonymized data trace information on a regular basis.
- Organize stakeholders in the area of advanced Internet traffic monitoring.
- Dissemination of project Results.
Project Overview

- 24 months durations
  - “officially” 27 months duration
- Started **Jan 2005**, scheduled to finish **December 2006**
  - “officially” started Oct. 2004
- Total budget = EUR 2.05 m, EC Contribution = EUR 1.625 m
- 9 partners:
  - FORTH, Greece (coordinator)
  - VU, The Netherlands
  - CESNET, Czech Republic
  - Uninett, Norway
  - ENDACE, UK
  - ALCATEL, France
  - FORTHnet, Greece
  - TNO, The Netherlands
  - TERENA, The Netherlands
- 6 Work Packages
Potential LOBSTER applications: GRIDs

- **GRID Performance debugging**
  - GRID-enabled applications are highly dependent on network characteristics
    - Remote data access
    - Remote resource access (e.g. sensors, instruments)
    - Remote computing power
  - How can we perform complete diagnosis when applications are not working as expected?
    - The local LAN? the WAN? The remote LAN?
    - The local computer? The remote server? A middleware server?
The Challenge
Why is it difficult?

• Scientific Challenge
  – Massive amounts of data (petabytes)
    • But, more than other sciences?
      – Such as astronomy?
  – 40 Gbps network speeds
    • Use “effective” sampling?

• Social – Business Challenge
  – Unwilling ISPs
    • Record a day in the life of “my routers”? 
    • Why give out information to competitors?
  – Trust
  – Privacy issues
    • Who “owns” the data? Is there any personal information recorded?

• Engineering challenge
  – Skilled personnel
    • To deploy and manage monitoring equipment
It gets even more complicated

- Port 80 traffic
  - 2002: 5%
  - 2005: 30%
  - At the same time all the other known ports have practically vanished

- From the known ports
  - We practically see only port 80
  - Why?

- We don't know
  - Maybe lots of applications use port 80 to bypass firewall restrictions

- Roughly speaking, current Traffic Characterization applications tell us:
- It is either port 80 or a dynamic port