Internet Worm and Virus Protection for Very High-Speed Networks

John W. Lockwood
Professor of Computer Science and Engineering

lockwood@arl.wustl.edu
http://www.arl.wustl.edu/~lockwood

Research Sponsor:

http://www.globalvelocity.info/
Internet Worms and Viruses

• The problem with worms and virus attacks
  – Annoyance to users
  – Costly to businesses (lost productivity)
  – Security threat to government (compromised data)

• Recent Attacks
  – Nimda, Code Red, Slammer
  – MSBlast
    • Infected over 350,000 hosts in Aug. 16, 2003
  – SoBigF
    • Infected 1 million users in first 24 hours
    • Infected > 200 million in the first week
    • Caused an estimated $1 billion in damages to repair.

• Detectable by a Signature in Content
  – Pattern of bytes
  – Regular Expression
  – Morphable pattern
Challenges to Stopping Worm and Virus Attacks

- End-systems difficult to maintain
  - Operating systems become outdated
  - Users introduce new machines on network

- Internet contains several types of traffic
  - Web, file transfers, telnet
  - Data may appear anywhere in the packet

- Networks process High Speed Data
  - Multi Gigabit/second data transmission rates now commonplace in campus, corporate, and backbone networks
  - Peer-to-Peer protocols dominate current and future traffic
  - Need Real-time gathering
    - No latency can be tolerated
Virus/Worm/Data Spread in Unprotected Networks
Virus/Worm/Data Spread in Unprotected Networks
Virus/Worm/Data Spread in Unprotected Networks
Virus/Worm/Data Spread in Unprotected Networks
Virus/Worm/Data Containment in Protected Networks

Content Scanning and Protection Device
Content Scanning Technology

- Fiber optic Line Cards
  - Gigabit Ethernet
  - ATM OC-3 to OC-48

- Reconfigurable Hardware
  - Uses Field Programmable Port Extender (FPX) Platform
  - Protocol processing and content scanning performed in hardware
  - Reconfigurable over the network

- Chassis / Motherboard
  - Allows Modules to Stack
Field-programmable Port Extender (FPX)
Remotely reprogramming hardware over the network

- New module developed
- Content Matching Server generates New module in programmable Logic
- Module Bitfile transmitted over network
- New module deployed into FPX hardware
Data Scanning Technologies

• Protocol Processing
  – Layered Protocol Wrappers
  – Process Cells/frames/packets/flows in hardware

• Regular Expression Matching
  – Deterministic Finite Automata (DFA)
  – Dynamically programmed into FPGA logic

• Fixed String Matching
  – Bloom Filters
  – Dynamically programmed into BlockRAMs
Regular Expression Matching with Finite Automata

Moscola et al.
String Matching with Bloom Filters

Dharmapulikar et al.
Complete Protection System

Network Aggregation Point (NAP)

Switch/Concentrator

Global Velocity DED

Router/Switch

Data

Regional Transaction Processor (RTP)

Data

Content Matching Server (CMS)/Central Storage and Backup System (CSBS)
System Components

• Hardware-based Data Processing
  – FPGA bitfile transferred over network to reconfigurable hardware
  – Content scanned in hardware with parallel Finite State Machines (FSMs)
  – Control messages sent over network allow blocking/unblocking of data

• Software-based System Generation
  – Web-based control and configuration
  – SQL Database stores signature patterns
  – Finite State Machines created with JLEX
  – VHDL-specified circuits generated, Instantiated, and integrated with Internet protocol processing wrappers
Selecting the Search Strings

<table>
<thead>
<tr>
<th>Select</th>
<th>Edit</th>
<th>Delete</th>
<th>Id</th>
<th>Search String</th>
<th>Description</th>
<th>Author</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>17</td>
<td>IHEX(6c744e5076)</td>
<td>Clear and Present Danger</td>
<td>9</td>
<td>3.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>6</td>
<td>ViRuS</td>
<td>An Email Virus</td>
<td>15</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>13</td>
<td>Copyright.* WashU</td>
<td>WashU Copyright</td>
<td>12</td>
<td>1.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>128</td>
<td>(L)(A)(D)(E)(N)</td>
<td>Terrorist Last Name</td>
<td>5</td>
<td>100.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>127</td>
<td>(O)j()sama</td>
<td>Terrorist First Name</td>
<td>5</td>
<td>5.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>112</td>
<td>Patient (Confidential</td>
<td>Record)</td>
<td>Confidential Information</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>113</td>
<td>Medical (Information</td>
<td>Record)</td>
<td>Medical Record</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>114</td>
<td>Do Not (Distribute</td>
<td>Release)</td>
<td>Confidential Information</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>129</td>
<td>IHEX(1B688E6D)</td>
<td>Internet Worm</td>
<td>19</td>
<td>6.00</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>130</td>
<td>NASA (C)(c) (confidential</td>
<td>CONFIDENTIAL)</td>
<td>Confidential Information</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>EDIT</td>
<td>DELETE</td>
<td>133</td>
<td>IHEX(683063423739)</td>
<td>SoBigF Internet Worm (MIME64)</td>
<td>16</td>
<td>11.00</td>
</tr>
</tbody>
</table>
Edit Search strings

Manage DED Library
Click "ADD" to generate a new entry.

search_string: IHEX(683063423739)
description: SoBigF Internet Worm (MIME64)
Author: 16
Value: 11.00

Update Entry
Program the Hardware

Click “Program DED Now” below to modify the predefined hardware devices. This process can take about 10 minutes.
Modular Design Flow
(our contribution)

Front End:
Specify Regular Expression
(Web, PHP)

Back End (1):
Extract Search terms from SQL database

Back End (2):
Generate Finite State Machines in VHDL

Synthesize Logic to gates & flops
(Synplicity Pro)

In-System, Data Scanning
on FPX Platform

Install and deploy modules over Internet
to remote scanners (NCHARGE)

Generate bitstream (Xilinx)

Place and Route with constraints (Xilinx)

Set Boundry I/O & Routing Constraints (DHP)

New, 2 Million-gate Packet Scanner:
9 Minutes
Network Configuration with Gigabit Ethernet

Data Enabling Device (DED) with FPX Processing Modules

Internet

Gigabit Ethernet

PC

PC

PC

PC
Passive Virus Protection

- **Internet User** requests information from Internet.
- Content returns from Internet through FPX.
- Content is processed in the FPX.
- Content containing virus is forwarded from FPX.
- Alert packet is sent to user to let them know of the virus.
- Internet User requests information from Internet.

In the diagram, a message is shown: "The message you are downloading may contain a virus. To ensure the protection of your system, you should use caution when viewing the message."
Active Virus Protection

Content requested from public Internet

Content returns from infected host

Content is processed in the FPX

Content containing virus is dropped at FPX

Alert packet is sent to user to let them know of the virus

(1) Data requested from public Internet

Internet User

Virus Agent

The message you are attempting to download contains a virus and has been halted.

To ensure the protection of your system, you should click the stop button on your browser and click OK.
Active Virus Example
Other Applications

• Prevent unauthorized release of data
  – Secure Classified documents
  – Lock medical documents for Health Insurance Portability and Accountability Act (HIPAA)

• Avoid liability for misuse of network
  – Copyright infringement
  – Pornography in the workplace
Content Scanning Technologies

- **General Purpose Microprocessors**
  - Fully Reprogrammable
  - Sequential Processing

- **Custom Packet Processing Hardware**
  - Highly concurrent processing
  - Static Functionality

- **Network Processors**
  - Mostly Reprogrammable
  - Some concurrent processing (8-32 cores)

- **Reconfigurable Hardware**
  - Fully Programmable
  - Highly concurrent processing
Performance

Throughput

Probability of Matching

FPGA-based Regular Expression Matching with Parallel Engines

Software-based Regular Expression Matching Systems (Snort, etc)
Actual Software Performance

Top Layer Networks & Internet Security Systems
Probability of Detection vs Percent Utilization

Throughput Comparison

• Sed was run on different Linux PCs
  – Dual Intel Pentium III @ 1 GHz
    • 13.7 Mbps when data is read from disk
    • 32.72 Mbps when data is read from memory
  – Alpha 21364 @ 667 MHz
    • 36 Mbps when data is read from disk
    • 50.4 Mbps when data is read from memory

• Software results are 40x slower than FPsed
Results

- Content Scanning Platform Implemented
  - Scans Internet packets for virus or Internet worm signatures using reconfigurable hardware
  - Generates prompts when matching content is found

- Content Matching Server Implemented
  - Automatically generates FPGA from regular expressions selected from database

- Regional Transaction Processor implemented
  - Tracks propagation of Internet worms and viruses

- Reduces the spread of malware from months to minutes
Acknowledgements

• Washington University
  – Faculty
    • John Lockwood
    • Ronald Loui
    • Jon Turner
  – Graduate Students
    • Mike Attig
    • Sarang Dharmapurikar
    • David Lim
    • Jing Lu
    • Bharath Madhusudan
    • James Moscola
    • Chris Neely
    • David Schuehler
    • Todd Sproull
    • David Taylor
    • Haoyu Song
    • Chris Zuver

• Industry Research Partners
  – Matthew Kulig (Global Velocity)
  – David Reddick (Global Velocity)
  – Tim Brooks (Global Velocity)

• Government Partners
  – National Science Foundation

• Hardware Vendors
  – David Parlour (Xilinx)

• Visiting Faculty and Students
  – Edson Horta
  – Florian Braun
  – Carlos Macian