Internet Worm and Virus Protection for Very High-Speed Networks

John Lockwood
Washington University in Saint Louis
lockwood@arl.wustl.edu - (314) 935-4460 - http://www.arl.wustl.edu/~lockwood

Abstract
The security of the Internet can be improved using reconfigurable hardware. A platform has been implemented that actively scans and filters Internet traffic at multi-Gigabit/second rates using reconfigurable hardware. Modular components implemented in FPGA logic process packet headers and scan for signatures of malicious software (malware) carried in packet payloads. Additional FPGA circuits track the state of Transmission Control Protocol (TCP) flows. Regular Expressions and fixed-string scanning circuits are implemented in parallel hardware. Dynamic reconfiguration enables remote modules to be reconfigured to scan for new signatures. Network-wide protection is achieved by the deployment of multiple systems throughout the Internet.

Introduction
Computer viruses and Internet worms cause billions of dollars in lost productivity. Well-known Internet worms like Nimda, Code Red and Slammer contain strings of malicious code that can be detected as they flow through the network. By processing the content of Internet traffic in real-time, a computer virus or Internet worm can be detected and prevented from propagating. Our system scans the full payload of packets to route, block, and account for the content in the flow. One challenge in implementing the system was that the location of a signature in the packet payload was not deterministic--it could appear at any position within the traffic flow. Another challenge to implementing the system was that signatures could span multiple packets and be interleaved among multiple traffic flows. The paper will describe how these challenges were met and overcome.

Related Work
A common requirement for network intrusion detection and prevention systems is the requirement to search for predefined signatures in the packet payload. Since conventional software-based algorithms for deep packet inspection have not kept pace with high-speed networks, hardware-based solutions are desirable. Hence, important building blocks of these systems include fast signature matching and protocol processing circuits. Most systems in this class have a common requirement for string matching. For example, a media file can be characterized by the presence of a string of bytes (for the rest of the paper, a string is synonymous to a signature) and its transmission across a link can be monitored by looking for the presence of this string on the link.

Key Contribution
Our key contribution is to envision, design and develop a cohesive malware protection system that includes an FPGA-based network platform, Internet protocol processing circuits, content matching modules, and automated design tools to enable the implementation and timely updating of network security applications in reconfigurable hardware. The system allows for the immediate blocking of known viruses and may be rapidly reprogrammed to recognize and block new threats. These upgrades are system-driven, and are not dependant upon actions by the end users to assure that the protection remains up to date.

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **20 AUG 2004**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**
4. TITLE AND SUBTITLE **Internet Worm and Virus Protection for Very High-Speed Networks**		5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Washington University in Saint Louis**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001694, HPEC-6-Vol 1 ESC-TR-2003-081; High Performance Embedded Computing (HPEC) Workshop (7th)., The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**35**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

The system's foundation is the Field-programmable Port Extender (FPX), which is implemented with two FPGAs, five banks of memory and two high-speed (OC-48 rate) network interfaces. The network interfaces connect to one of several types of Gigabit-speed line card interface cards, including several types of Gigabit Ethernet and ATM interfaces. On the FPX, one FPGA is used to route individual traffic flows through the device, while the other is dynamically reconfigured over the network to perform customized packet processing functions. Using the latest FPGA technology, the system could easily scale to process 10 Gigabit/second OC-192 flows.

A TCP/IP wrapper, implemented in FPGA logic, reconstructs the flow of transmitted data by tracking sequence numbers of consecutive packets to provide a byte-ordered data stream to the content scanning engines. This means that even if a malware signature has been fragmented across multiple packets, it still will be detected and blocked. In order to maintain the state of multiple traffic flows, the system architecture has been designed to store the state of a TCP/IP flow in memory. Given that each flow occupies 64 bytes of memory, one 512 Mbyte SDRAM (about half of the memory on the FPX) module can track 8 million simultaneous traffic flows.

Two methods are used to search for signatures: a finite automata scans for regular expressions and a Bloom filter scans for fixed strings. The number of regular expressions that can be searched grows with the amount of the FPGA logic on the device, while the number of fixed strings that can be searched grow with the size of on-chip RAM. A Bloom filter allows a scanning engine to identify up to 1,700 fixed-length strings. Both types of our engines can scan traffic at traffic at 600 Mbps. By implementing four engines that run in parallel, the FPX can process data at a rate of 2.4 Gigabits per second using a single Xilinx Virtex 2000E FPGA.

An automated design flow builds packet scanning circuits in hardware. Custom circuits are built by an automated program that reads a list of signatures from a database table, optimizes each finite automata, integrates Internet protocol processing hardware, compiles the circuit into gates, routes and places the circuit into a FPGA, and then reconfigures remote devices over the network.

Conclusions

We have designed and developed a system that blocks the spread of Internet worms and computer viruses. Our system uses reconfigurable hardware to scan Internet traffic for malware. Malware is identified by signatures that may consist of either fixed strings or regular expressions. TCP/IP flows are tracked so that signatures spanning multiple packets can be detected. An automated design flow allows new circuits to be rapidly deployed to protect the network against new attacks.

References

- J. W. Lockwood. An open platform for development of Network processing modules in reprogrammable hardware. In IEC DesignCon'01, pages WB–19, Santa Clara, CA, Jan. 2001.
- R. Sidhu and V. K. Prasanna. Fast Regular Expression Matching using FPGAs. Field-Programmable Custom Computing Machines (FCCM), Rohnert Park, CA, Apr. 2001.
- R. Fanklin, D. Caraver, and B. Hutchings. Assisting network intrusion detection with reconfigurable hardware. Field Programmable Custom Computing Machines (FCCM), Apr. 2002.
- M. Fisk and G. Varghese. Fast content-based packet handling for intrusion detection. Technical Report CS2001-0670, University of California, San Diego, 2001.
- J. W. Lockwood, N. Naufel, J. S. Turner, and D. E. Taylor. Reprogrammable Network Packet Processing on the Field Programmable Port Extender (FPX). In ACM International Symposium on Field Programmable Gate Arrays (FPGA), pages 87–93, Monterey, CA, USA, Feb. 2001.
- J. Moscola, J. Lockwood, and R. P. Loui. Implementation of a Content-Scanning Module for an Internet Firewall. Field-Programmable Custom Computing Machines (FCCM), Apr. 2003.
- M. Necker, D. Contis, and D. Schimmel. TCP-Stream Poster on Reassembly and State Tracking in Hardware. Field-Programmable Custom Computing Machines (FCCM), Apr 2002.
- D. V. Schuehler and J. W. Lockwood. TCP-Splitter: A TCP/IP Flow Monitor in Reconfigurable Hardware. Symposium on High Performance Interconnects (HotI), pages 127–131, Stanford, CA, USA, Aug. 2002.

Internet Worm and Virus Protection for Very High-Speed Networks

John W. Lockwood
Professor of Computer Science and Engineering
Washington University in St. Louis
lockwood@arl.wustl.edu
http://www.arl.wustl.edu/~lockwood

Research Sponsor:
GLOBAL VELOCITY
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)

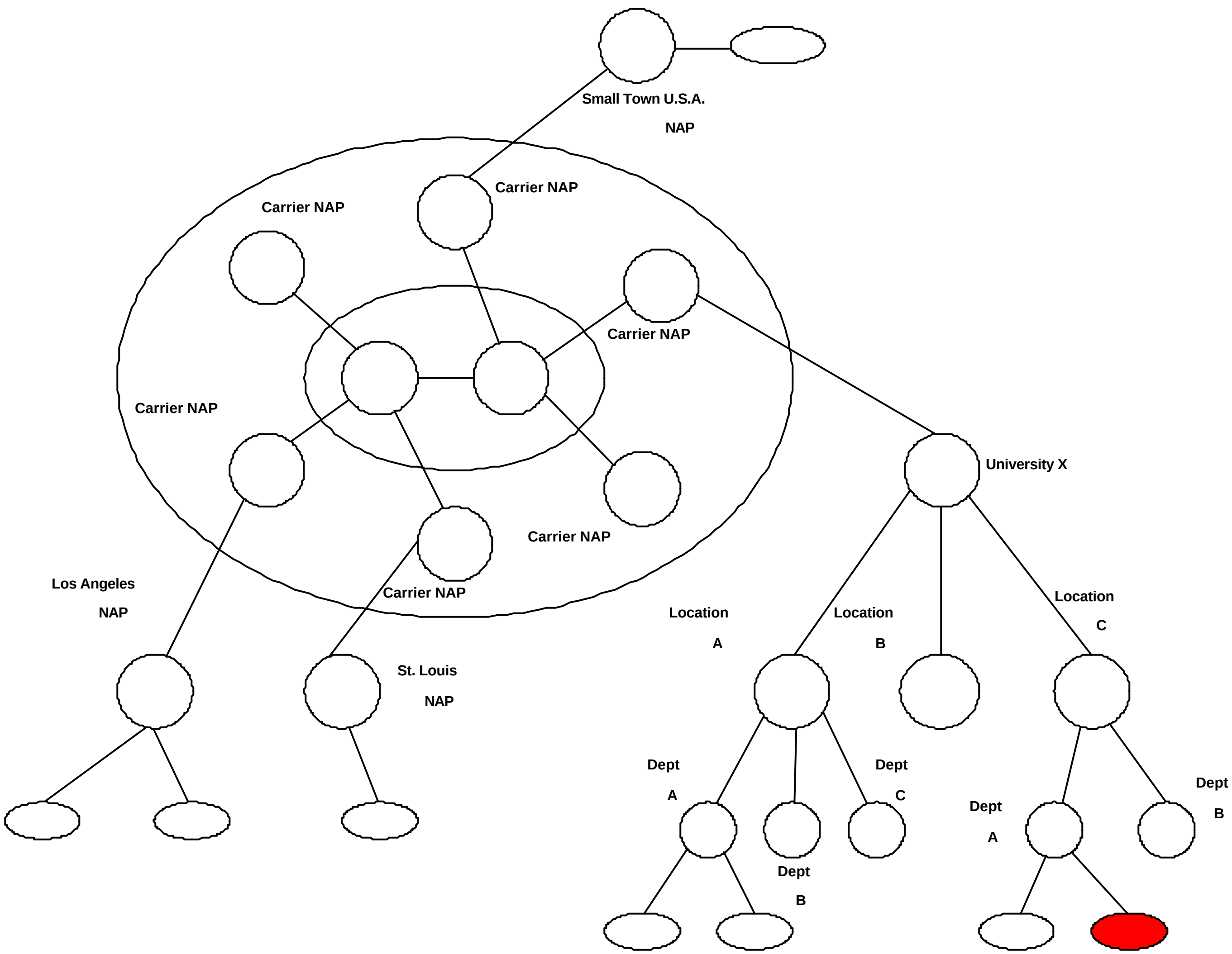
- Recent Attacks
  - Nimda, Code Red, Slammer
  - MSBlast
    - Infected over 350,000 hosts in Aug. 16, 2003
  - SoBigF
    - Infected 1 million users in first 24 hours
    - Infected > 200 million in the first week
    - Caused an estimated $1 billion in damages to repair.

- Detectable by a Signature in Content
  - Pattern of bytes
  - Regular Expression
  - Morphable pattern

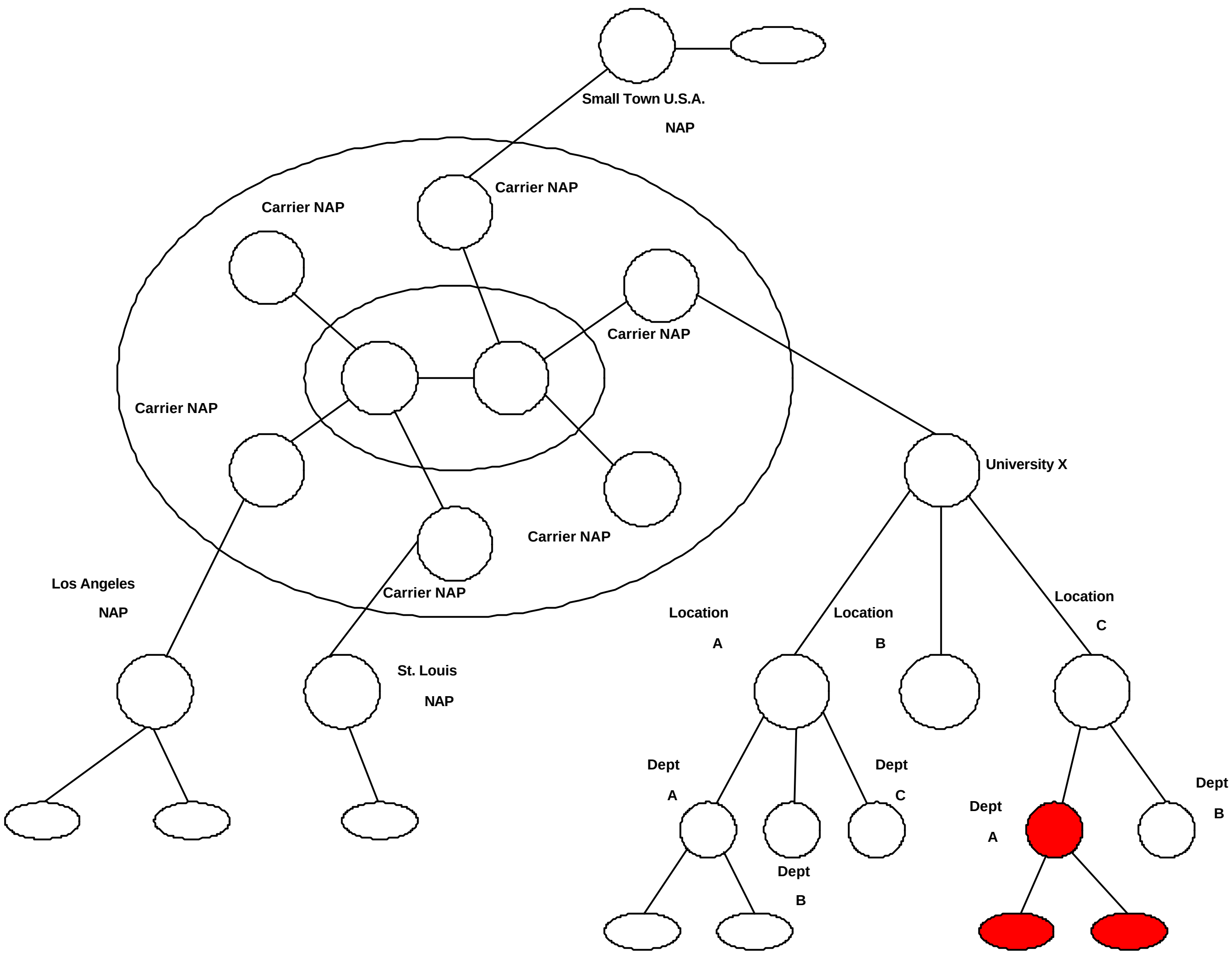
Challenges to Stopping Worm and Virus Attacks

- 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

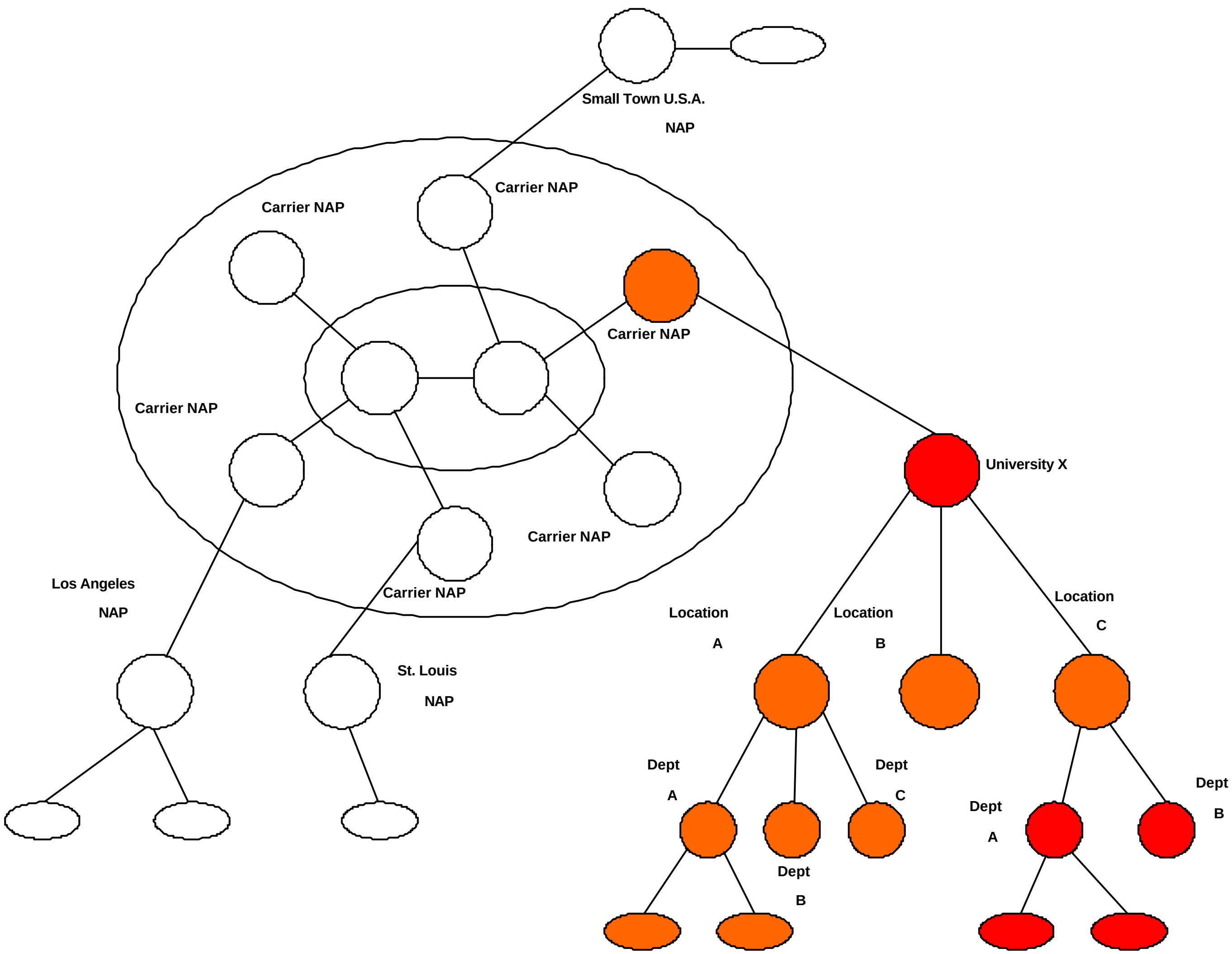


Virus/Worm/Data Spread in Unprotected Networks



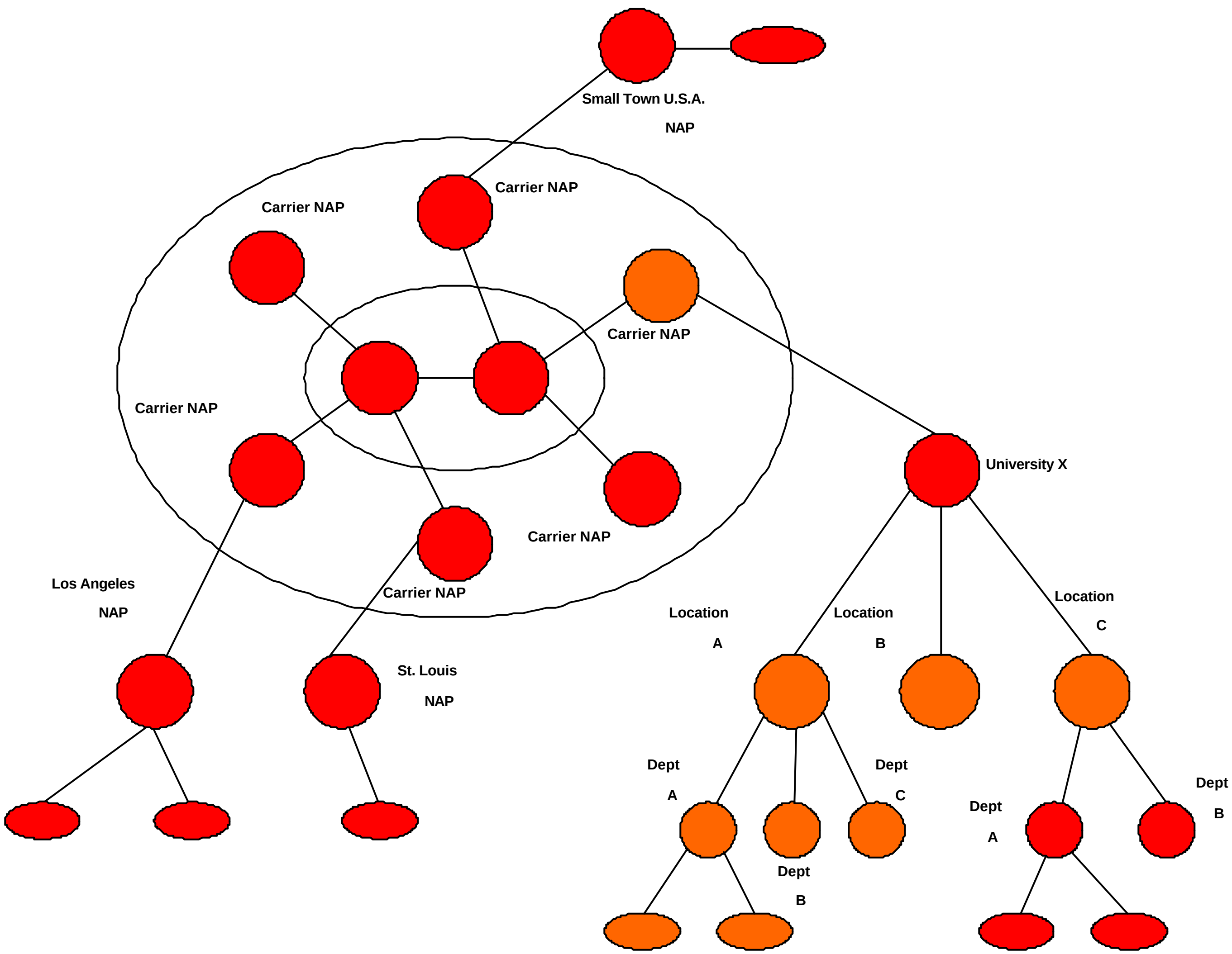
Washington University in St. Louis

Virus/Worm/Data Spread in Unprotected Networks

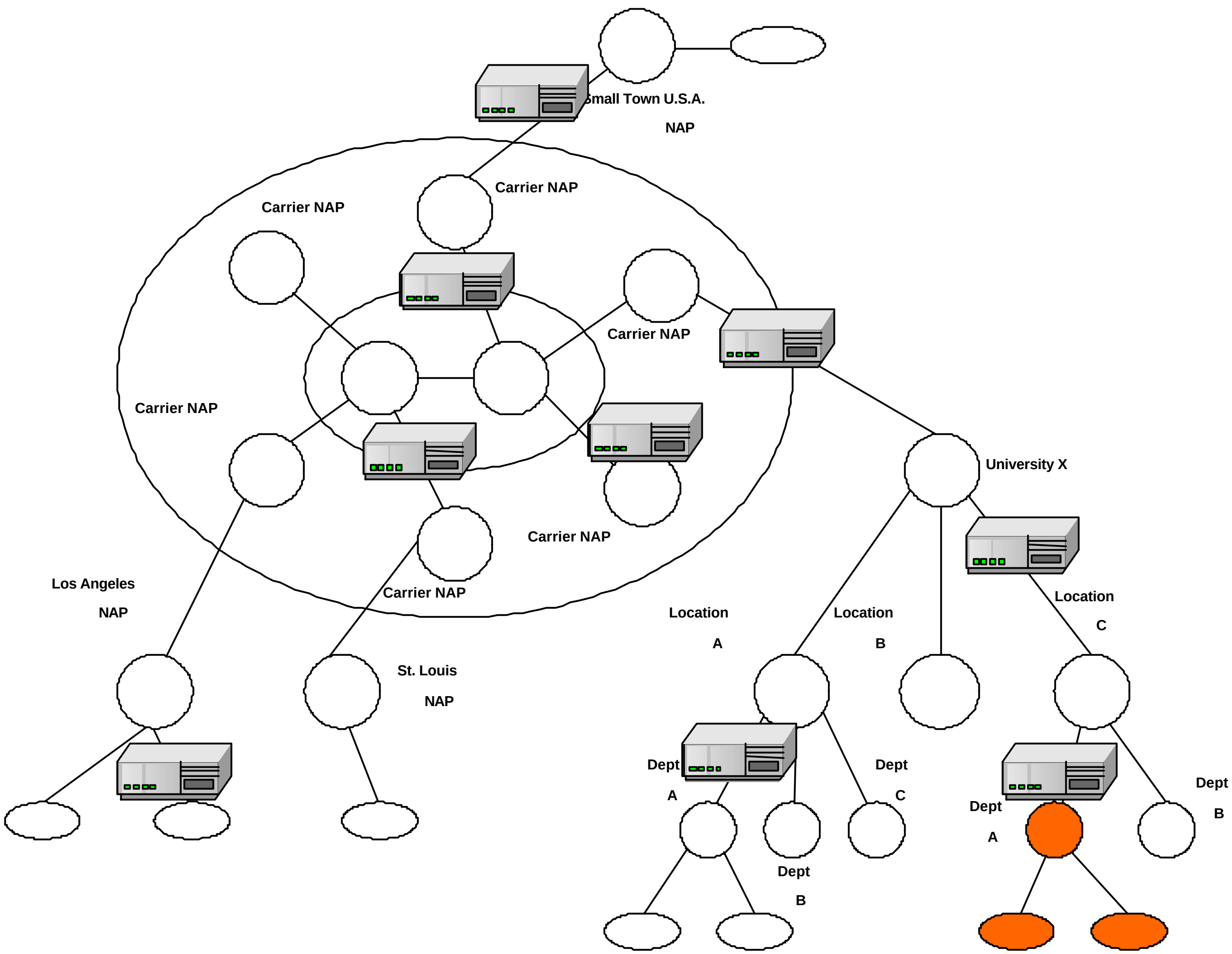


Washington University in St. Louis

Virus/Worm/Data Spread in Unprotected Networks

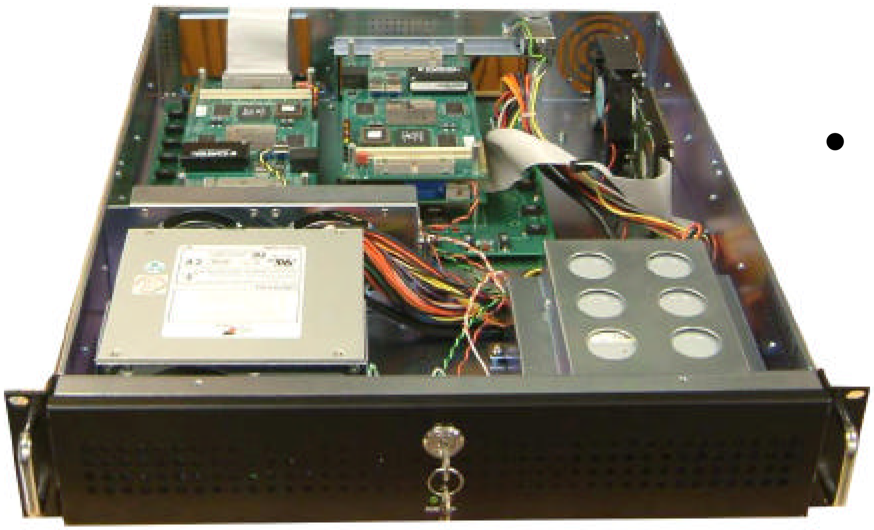


Virus/Worm/Data Containment in Protected Networks



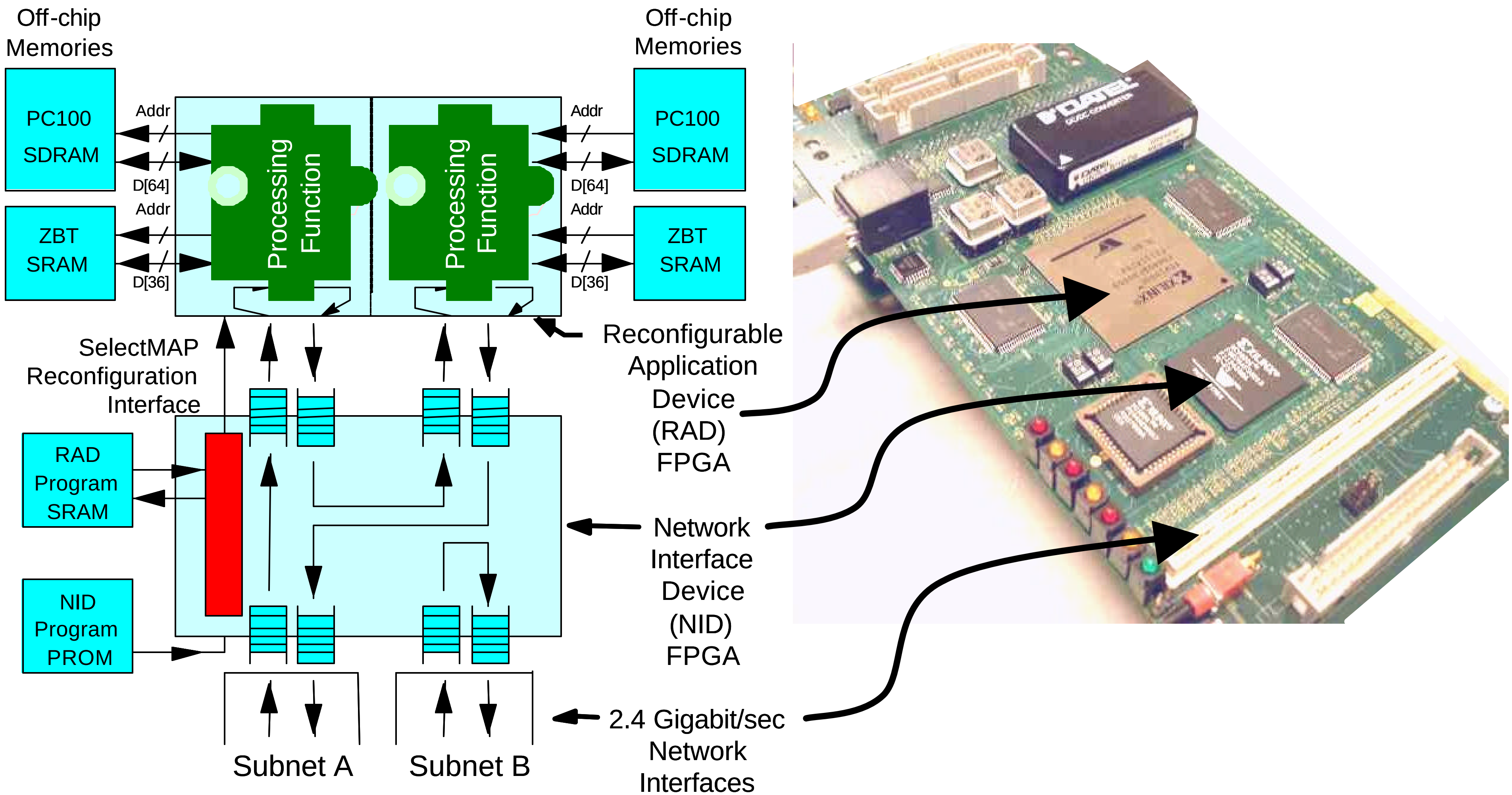
Content Scanning and Protection Device

Content Scanning Technology

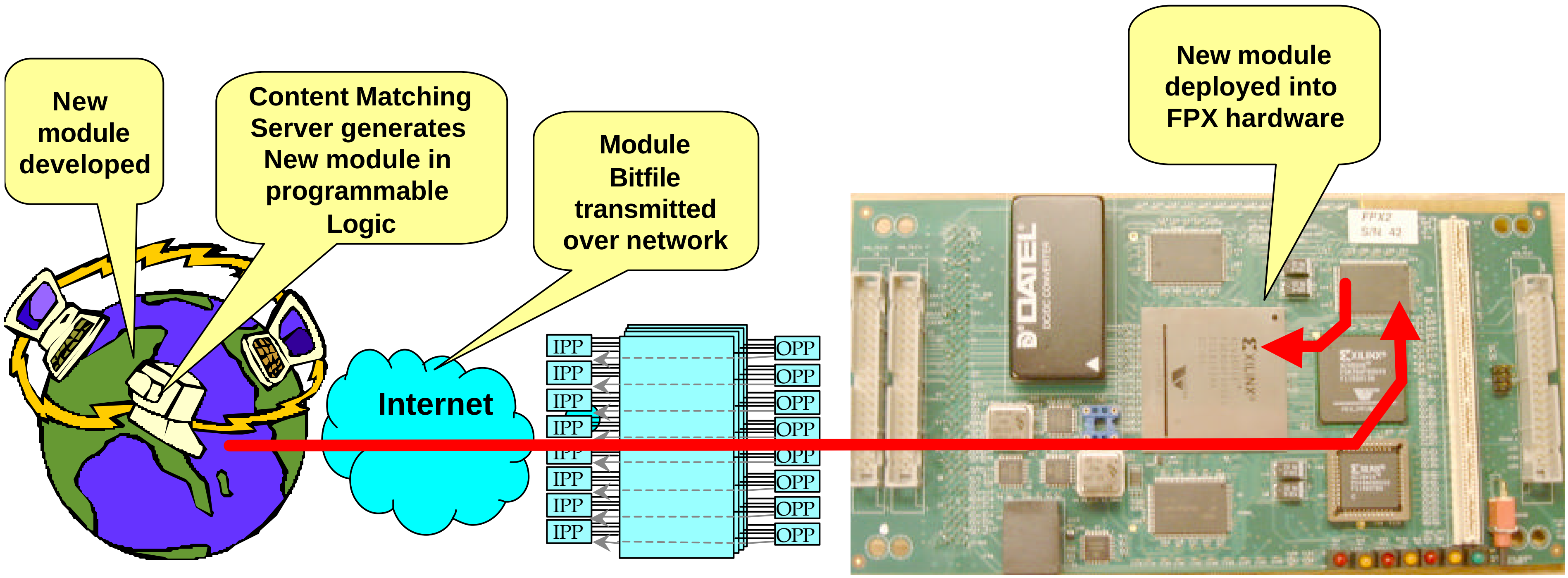


- Fiber optic Line Cards
  - Gigabit Ethernet
  - ATM OC-3 to OC-48
- Reconfigurable Hardware
  - Uses Field Programmable Port Extender (FPX) Platform
  - Protocol processing and content scanning performed in hardware
  - Reconfigurable over the network
- Chassis / Motherboard
  - Allows Modules to Stack

Field-programmable Port Extender (FPX)



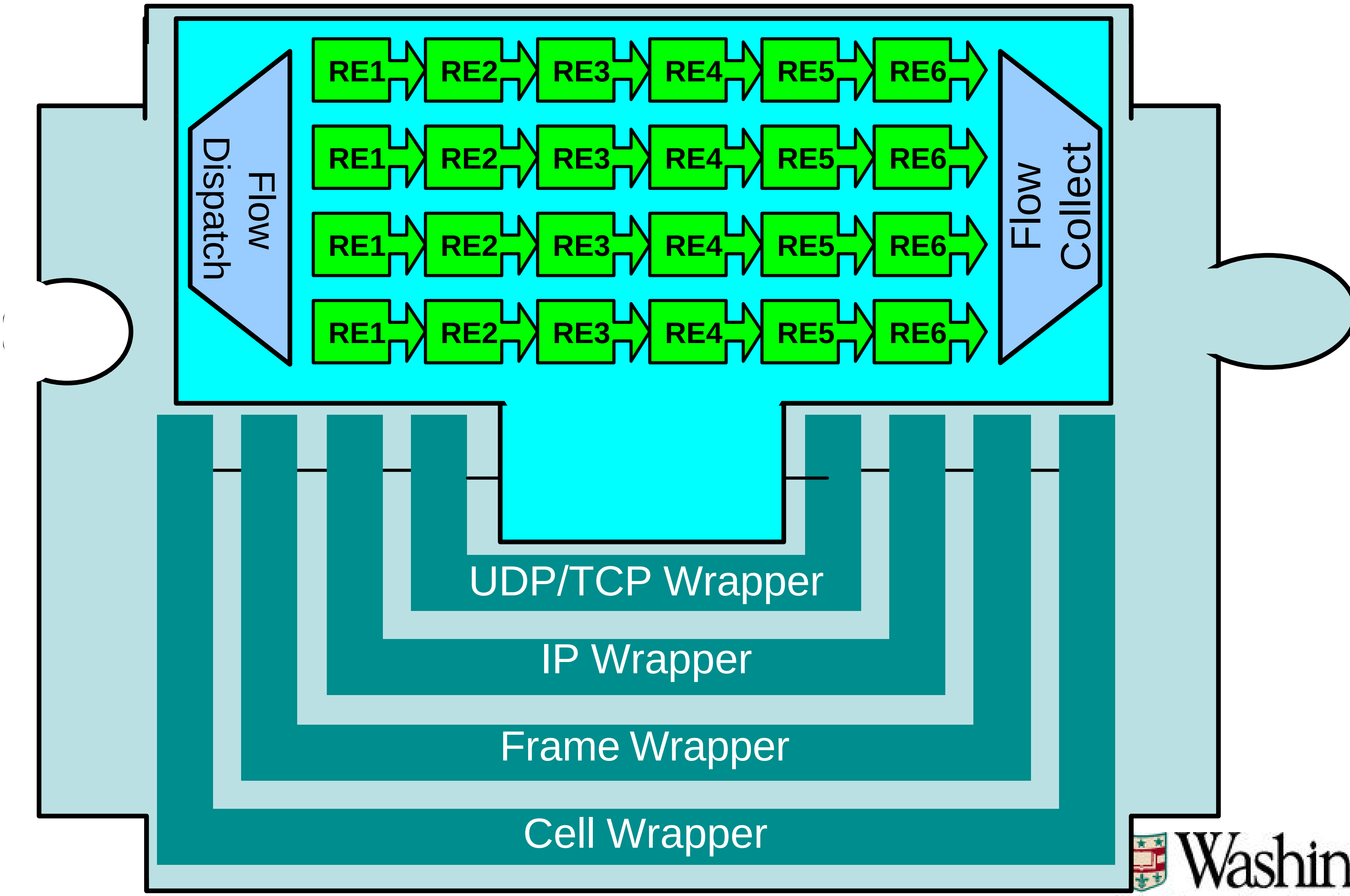
Remotely reprogramming hardware over the network



Data Scanning Technologies

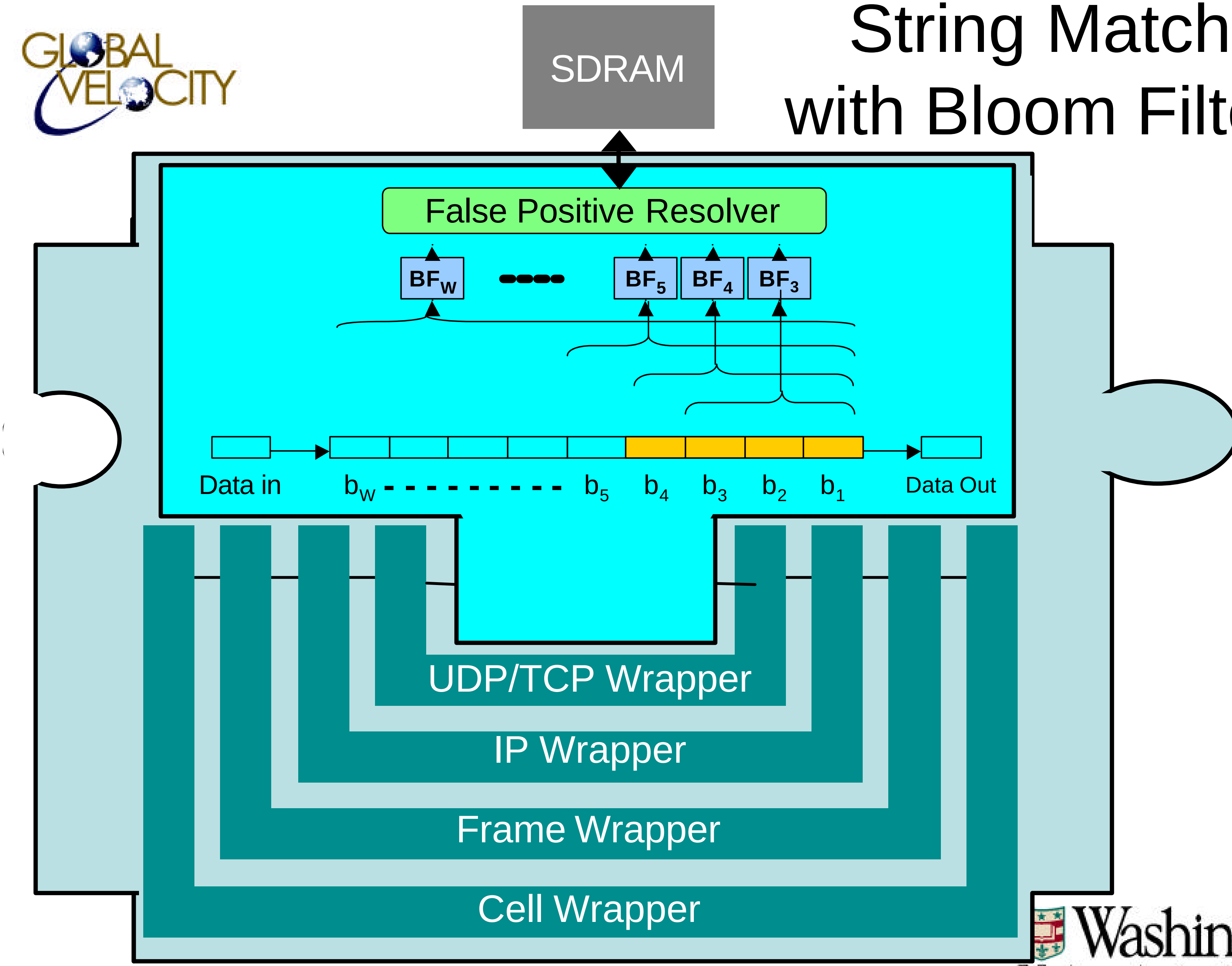
- Protocol Processing
  - Layered Protocol Wrappers
  - Process Cells/frames/packets/flows in hardware
- Regular Expression Matching
  - Deterministic Finite Automata (DFA)
  - Dynamically programmed into FPGA logic
- Fixed String Matching
  - Bloom Filters
  - Dynamically programmed into BlockRAMs

Regular Expression Matching with Finite Automata



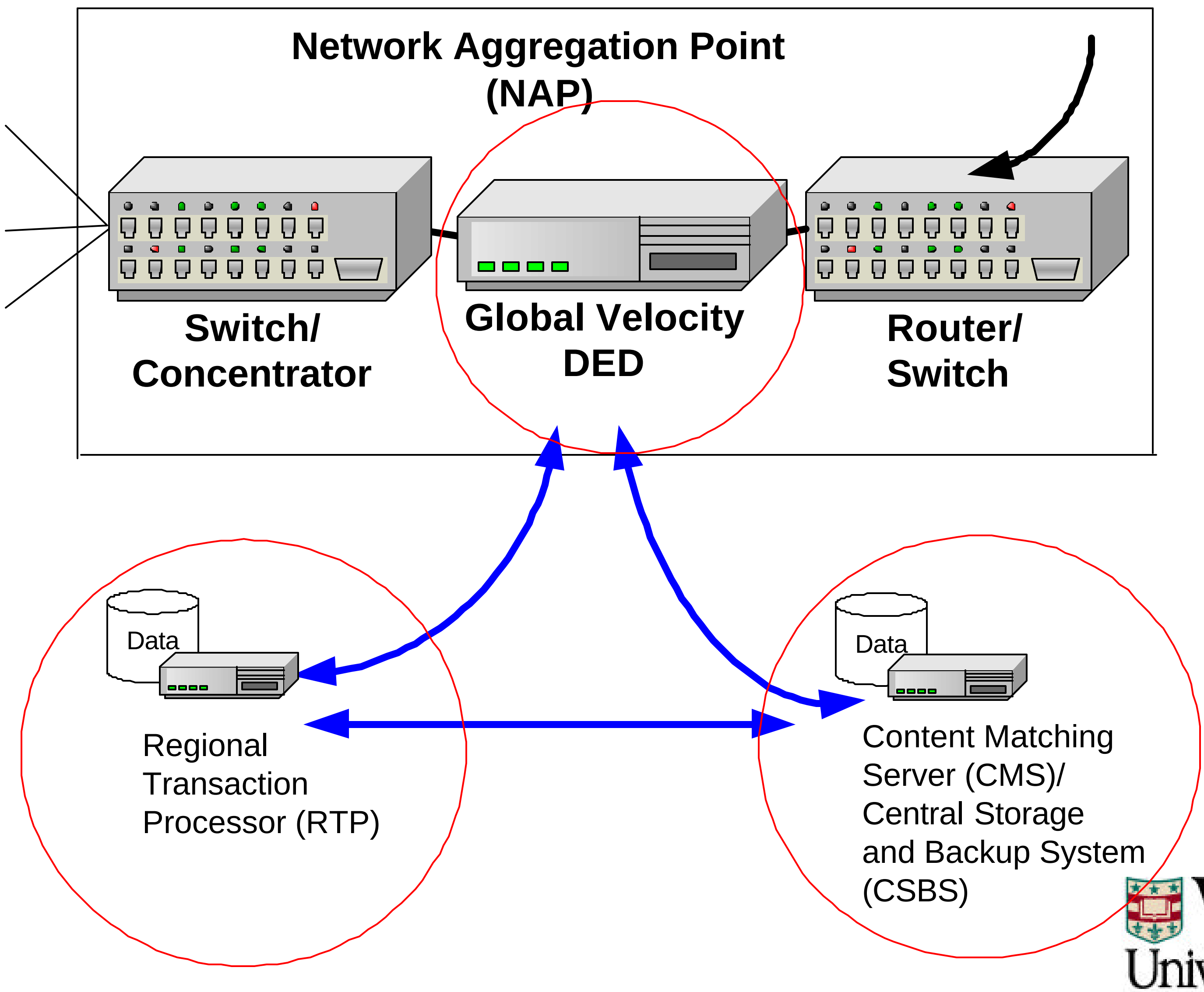
Moscola et al.

String Matching with Bloom Filters



Dharmapulikar et al.

Complete Protection System

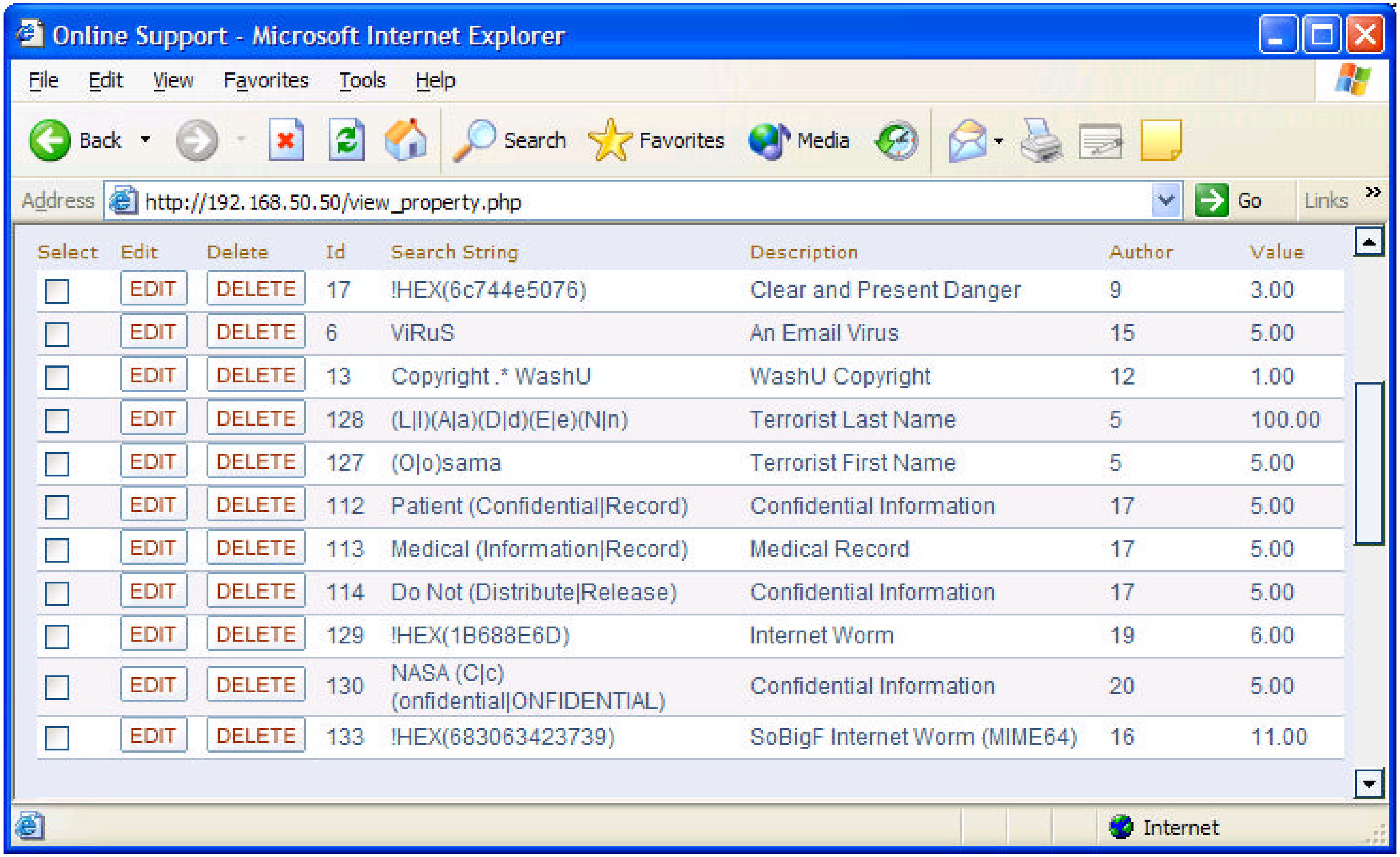


System Components

- 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


Online Support - Microsoft Internet Explorer

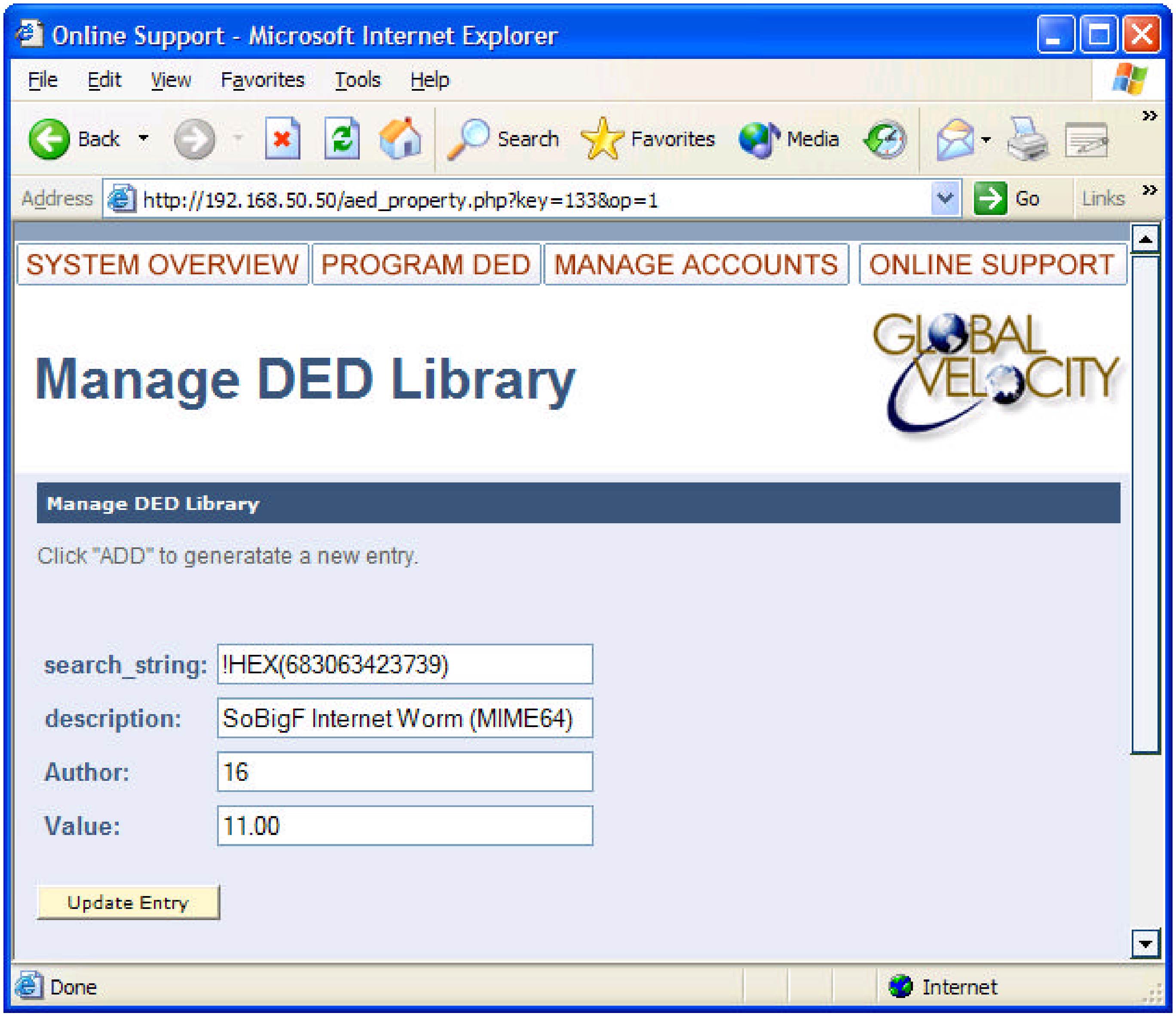
File Edit View Favorites Tools Help

Address http://192.168.50.50/view_property.php

Select	Edit	Delete	Id	Search String	Description	Author	Value
☐	EDIT	DELETE	17	!HEX(6c744e5076)	Clear and Present Danger	9	3.00
☐	EDIT	DELETE	6	ViRuS	An Email Virus	15	5.00
☐	EDIT	DELETE	13	Copyright .* WashU	WashU Copyright	12	1.00
☐	EDIT	DELETE	128	(L|l)(A|a)(D|d)(E|e)(N|n)	Terrorist Last Name	5	100.00
☐	EDIT	DELETE	127	(O|o)sama	Terrorist First Name	5	5.00
☐	EDIT	DELETE	112	Patient (Confidential|Record)	Confidential Information	17	5.00
☐	EDIT	DELETE	113	Medical (Information|Record)	Medical Record	17	5.00
☐	EDIT	DELETE	114	Do Not (Distribute|Release)	Confidential Information	17	5.00
☐	EDIT	DELETE	129	!HEX(1B688E6D)	Internet Worm	19	6.00
☐	EDIT	DELETE	130	NASA (C|c)(onfidential|ONFIDENTIAL)	Confidential Information	20	5.00
☐	EDIT	DELETE	133	!HEX(683063423739)	SoBigF Internet Worm (MIME64)	16	11.00

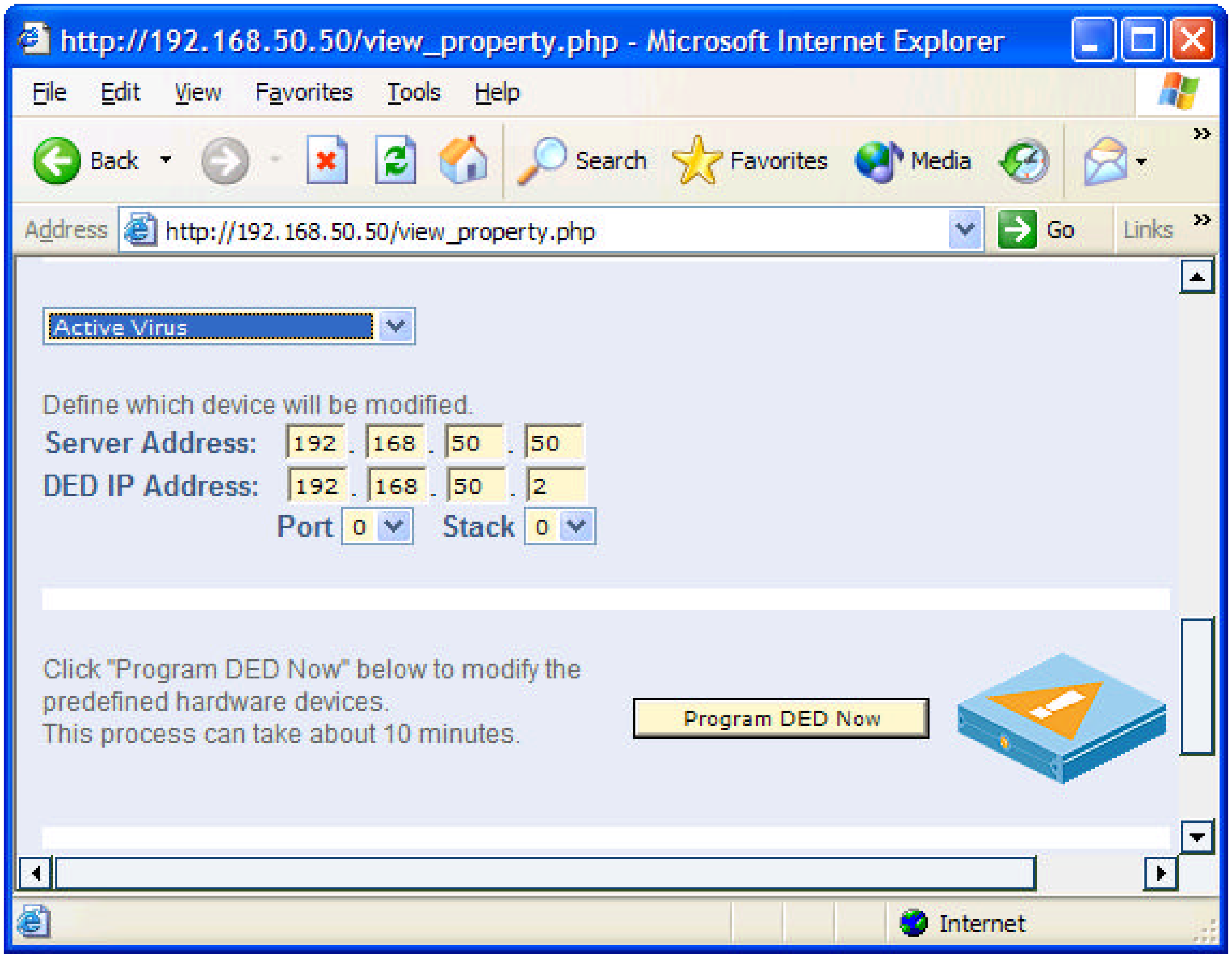
Internet

Edit Search strings



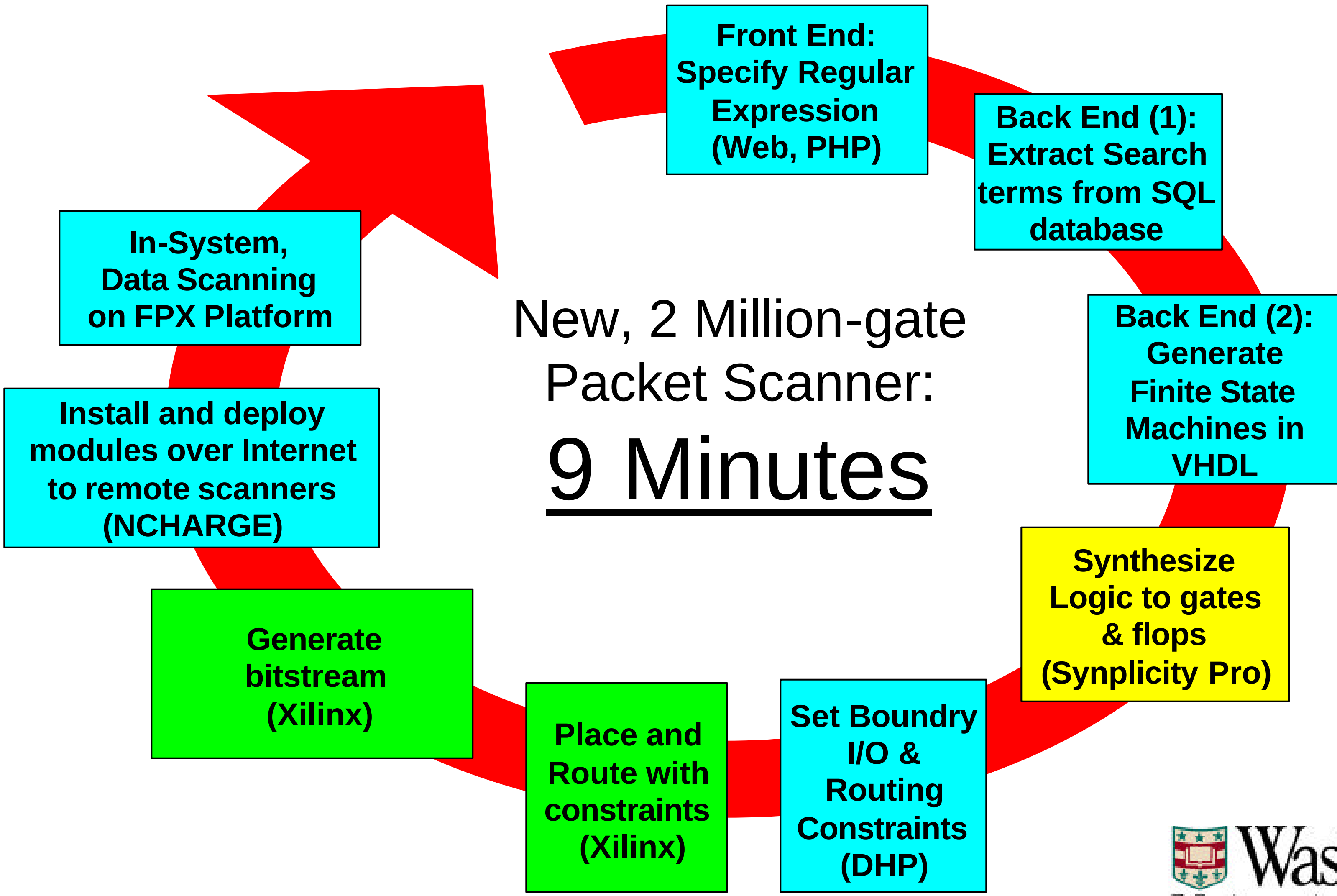
Washington University in St.Louis

Program the Hardware



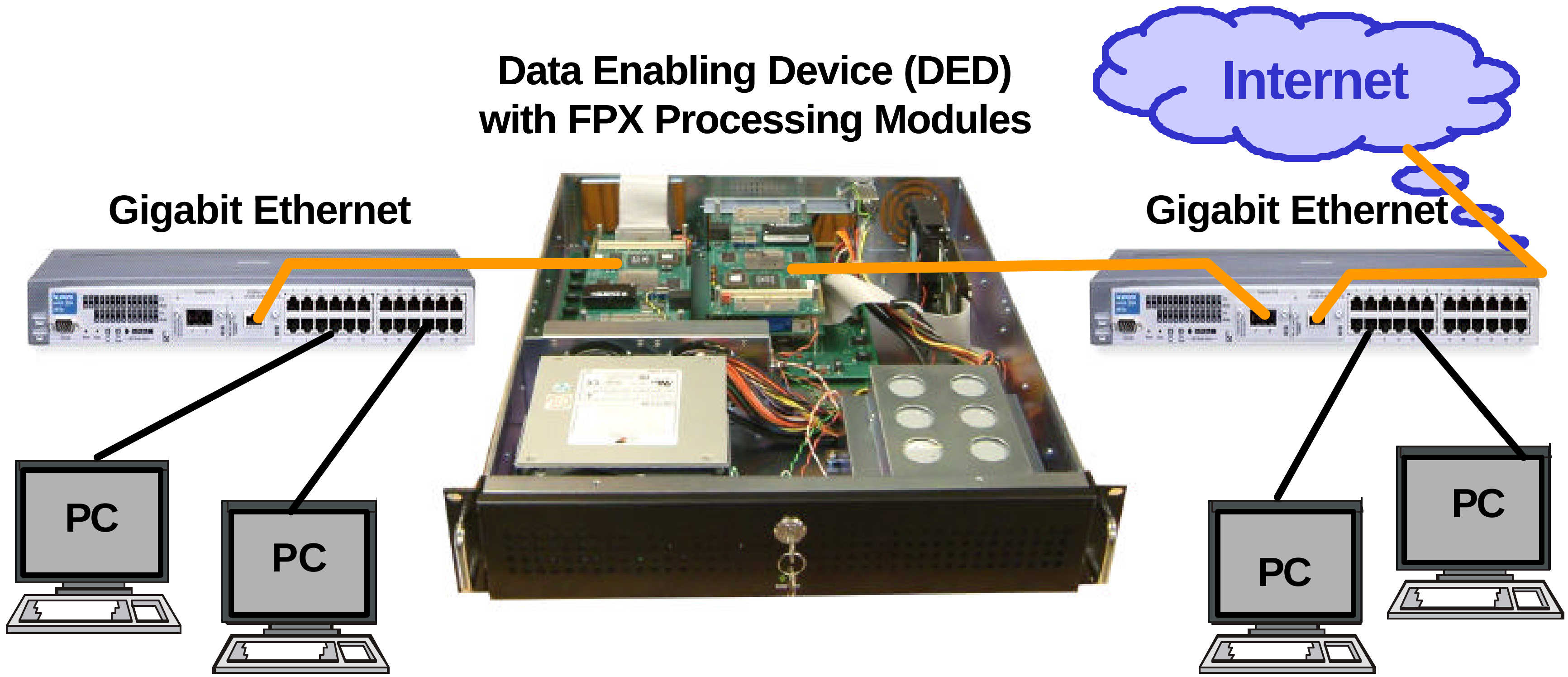
Modular Design Flow

(our contribution)

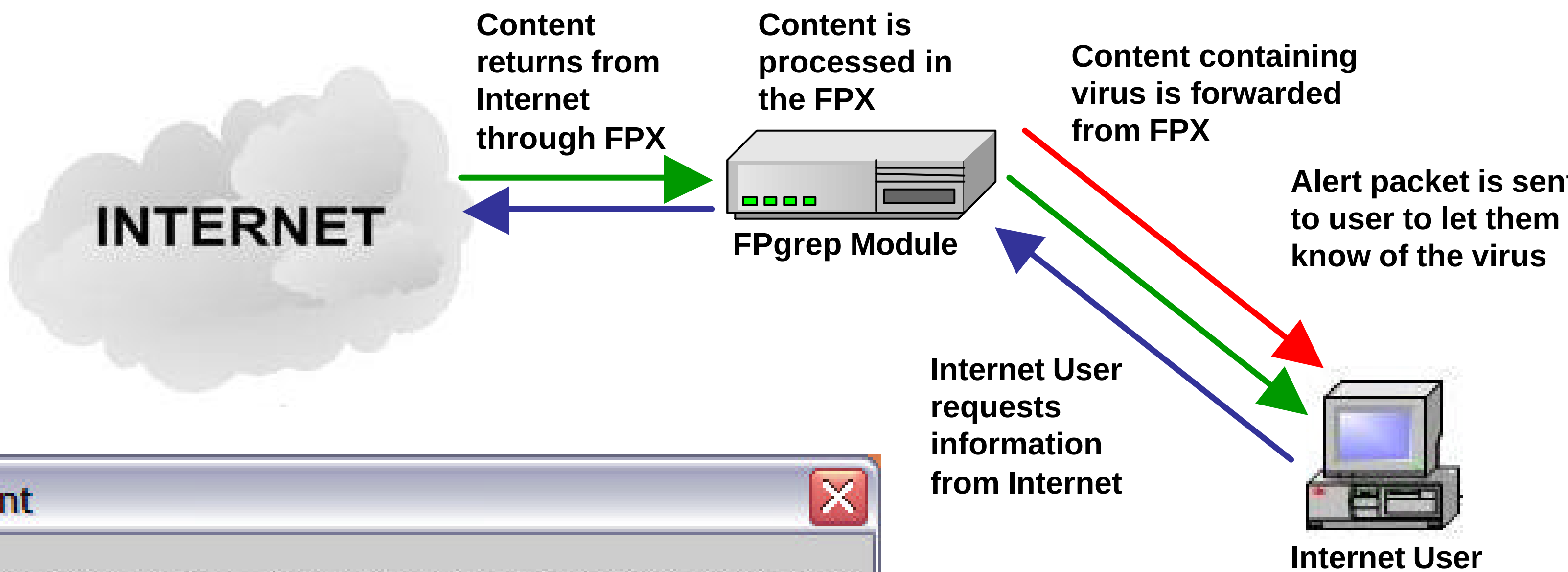


Washington University in St. Louis

Network Configuration with Gigabit Ethernet

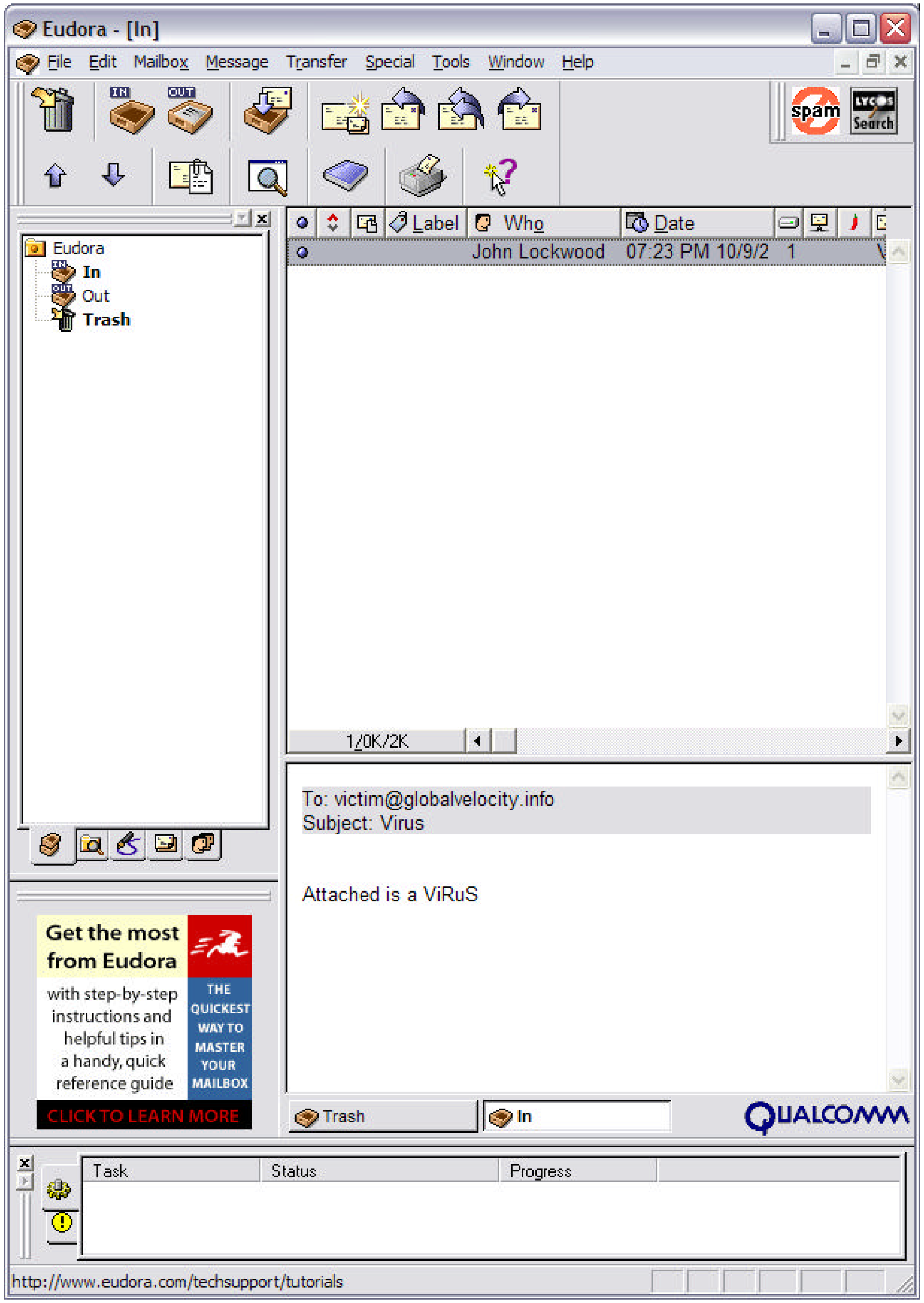


Passive Virus Protection

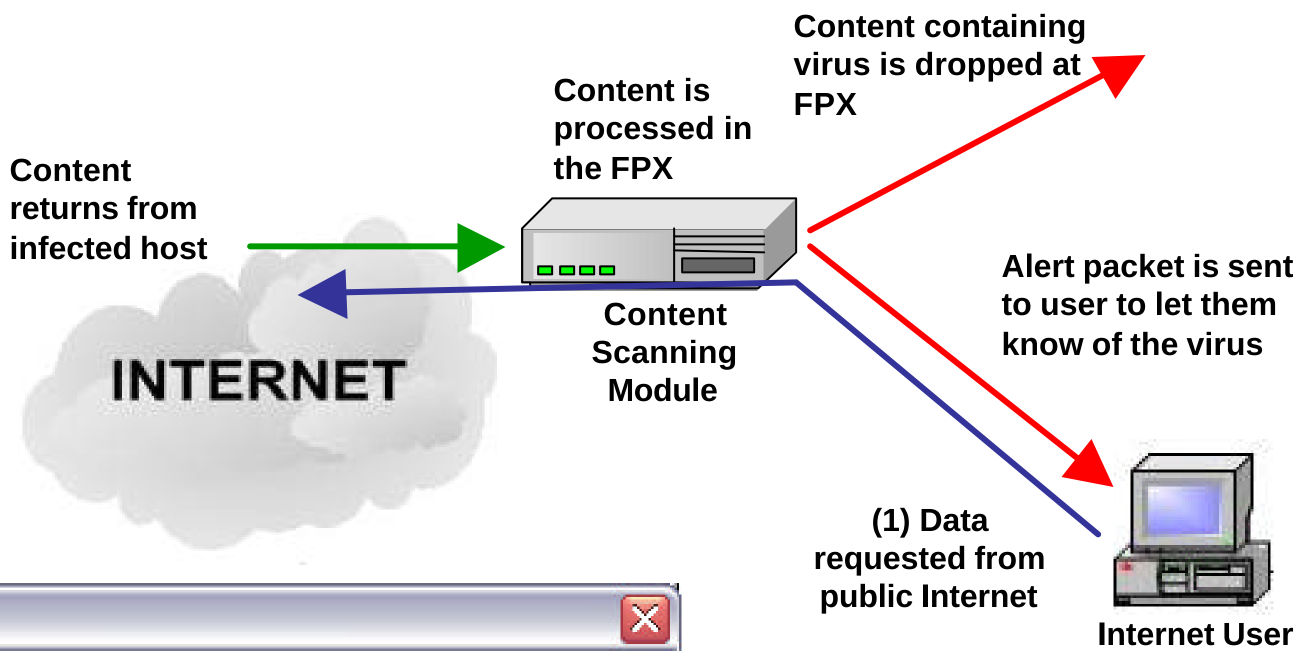


Washington University in St.Louis

Passive Virus Example

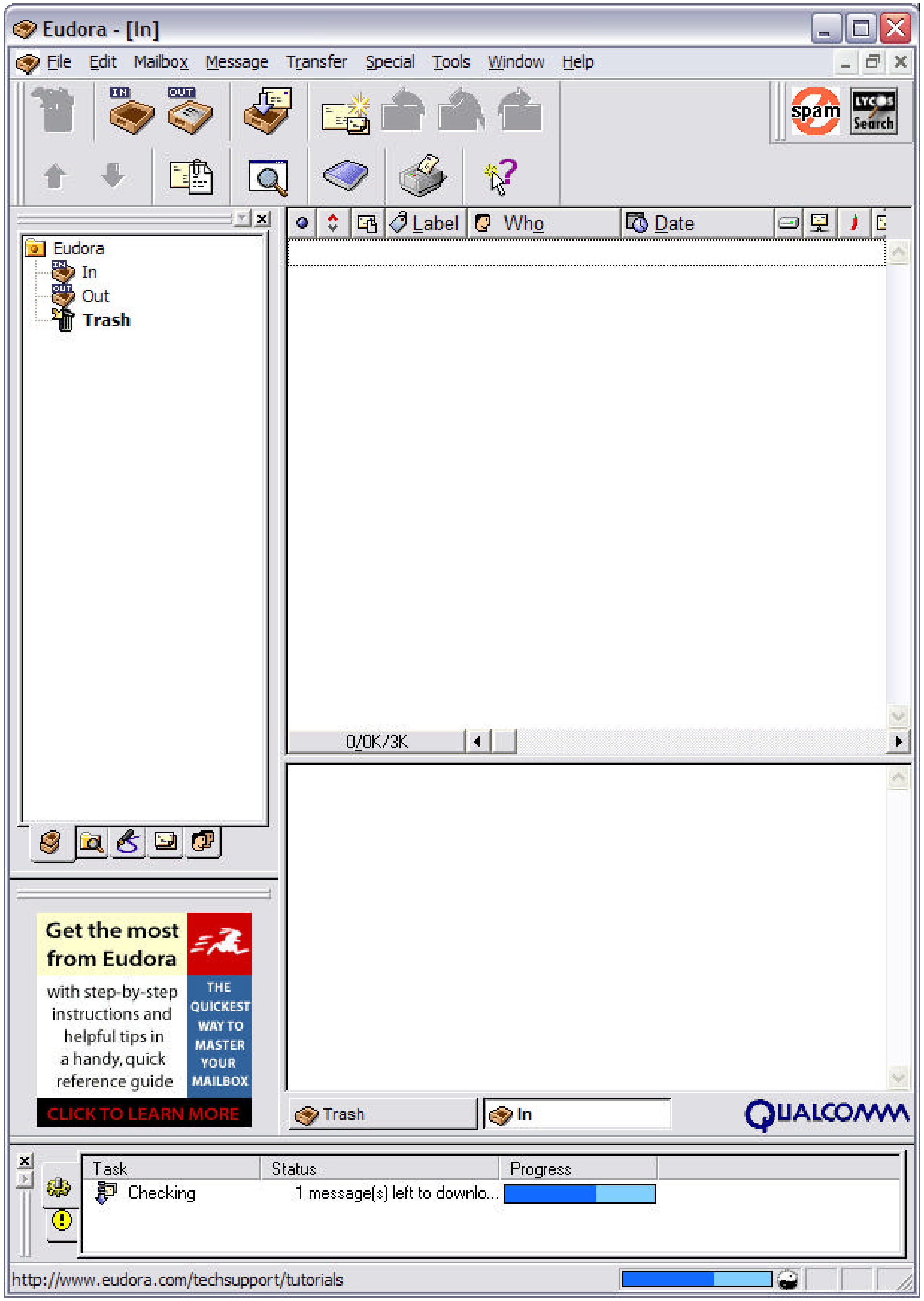


Active Virus Protection



Washington University in St. Louis

Active Virus Example



Washington University in St.Louis

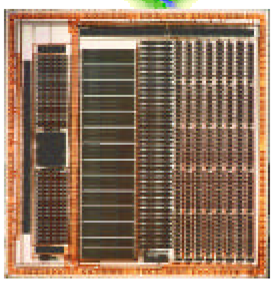
Other Applications

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

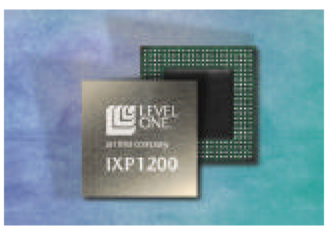
- Avoid liability for misuse of network
  - Copyright infringement
  - Pornography in the workplace

Content Scanning Technologies

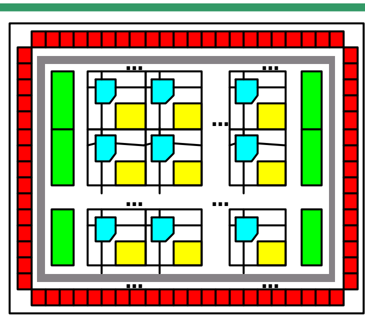
- General Purpose Microprocessors
  - Fully Reprogrammable
  - Sequential Processing



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



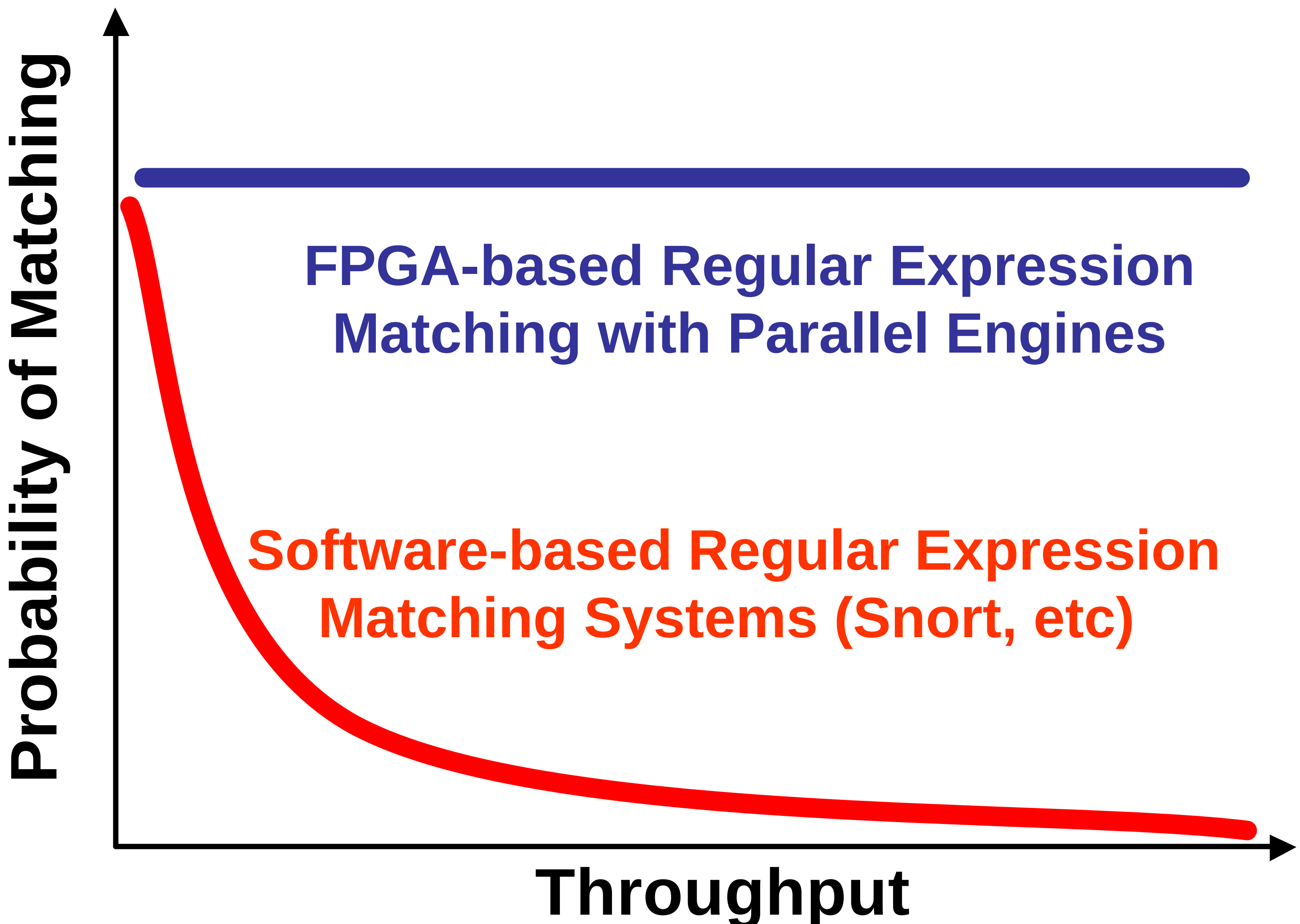
- Network Processors
  - Mostly Reprogrammable
  - Some concurrent processing (8-32 cores)



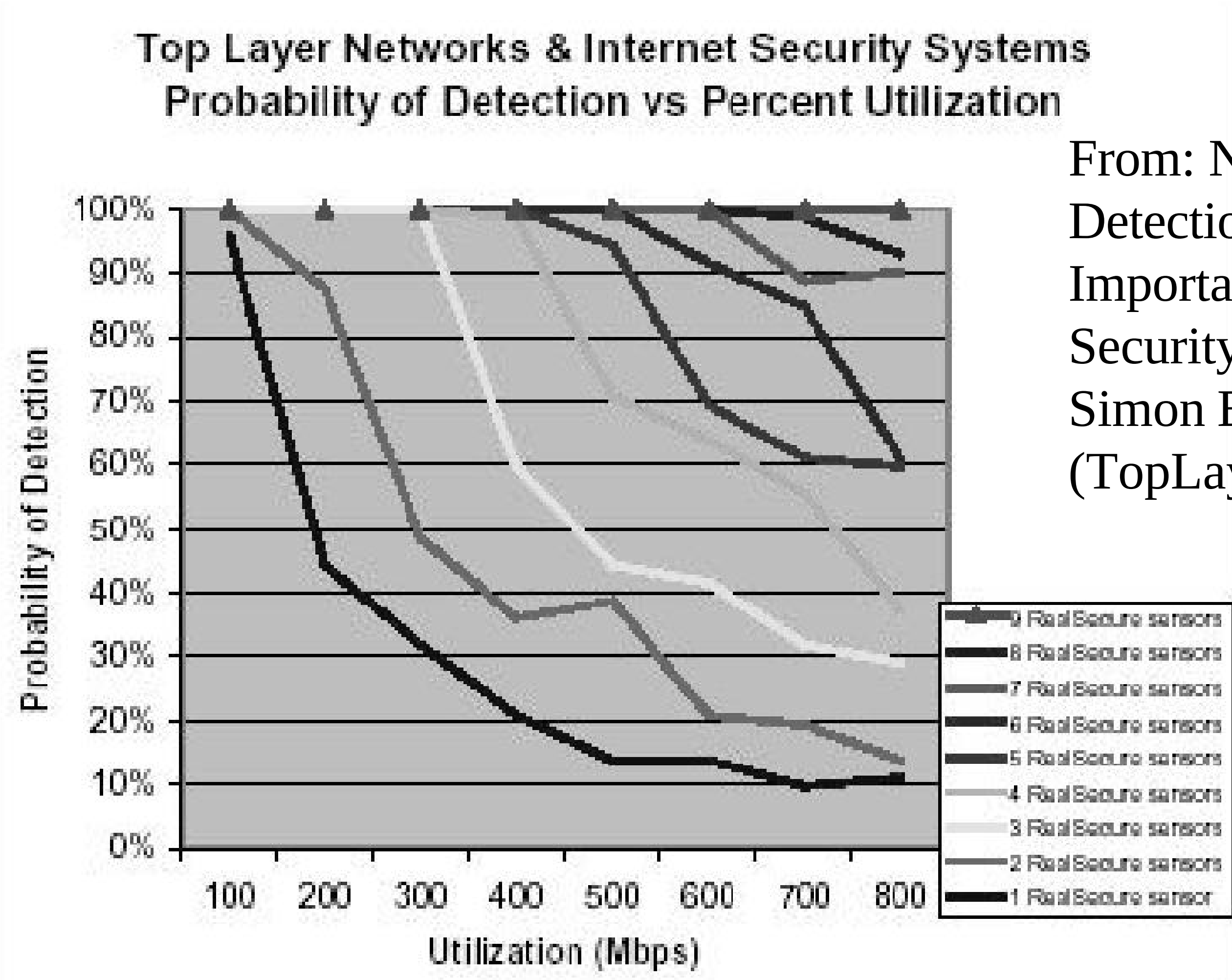
- Reconfigurable Hardware
  - Fully Programmable
  - Highly concurrent processing

Washington University in St. Louis

Performance



Actual Software Performance

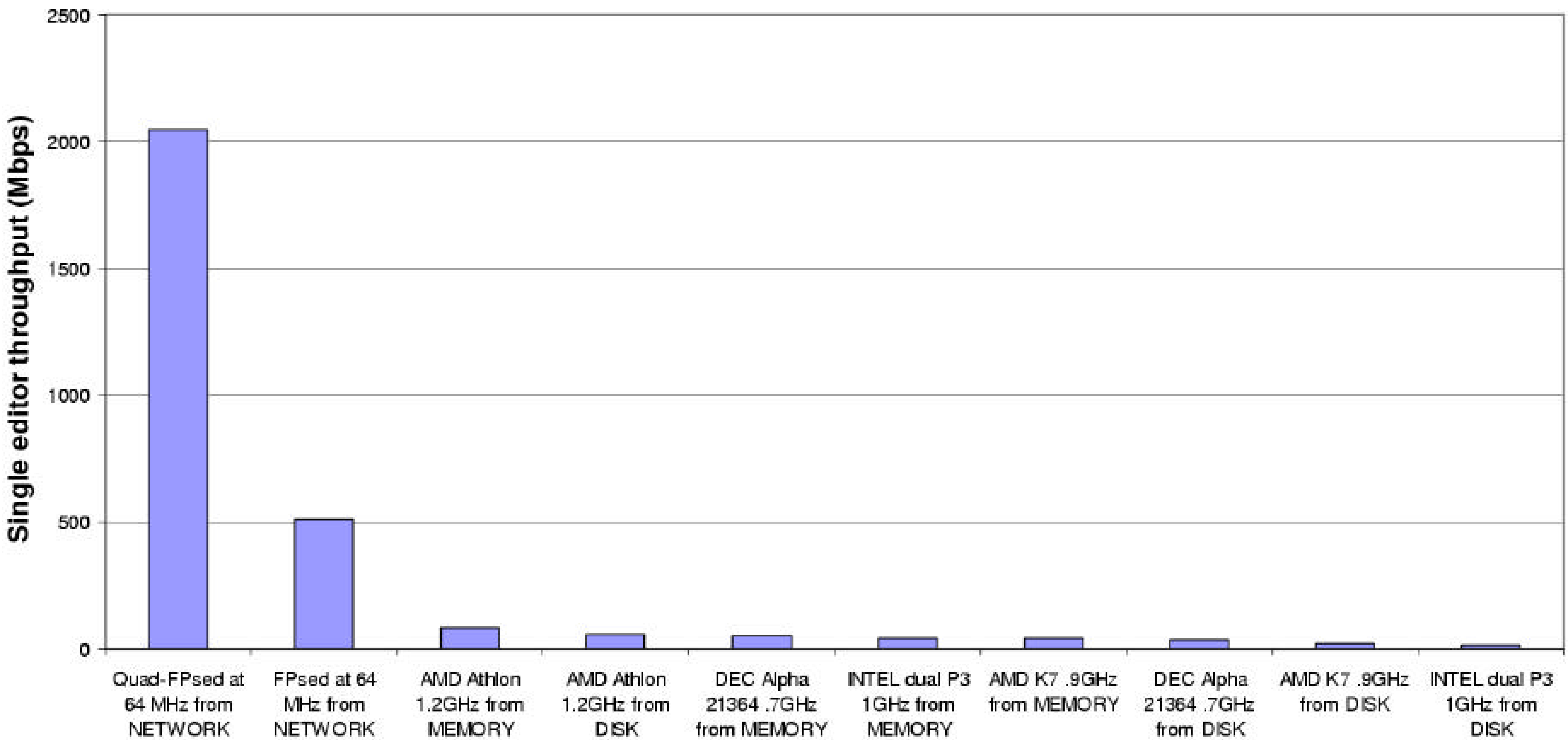


From: Network Intrusion Detection Systems: Important IDS Network Security Vulnerabilities by Simon Edwards (TopLayer.com)

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

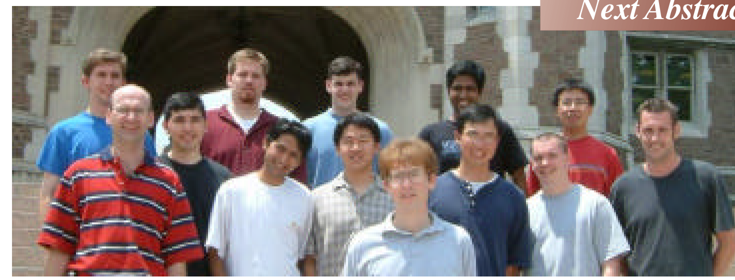
String Processing Benchmarks (measured results for SED)



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