Fidelity/Scalability Tradeoffs in Protocol Evaluation

Sonia Fahmy

DETER/EMIST work is joint with: Roman Chertov, Ness B. Shroff, and a group of B.S. and M.S. students

Center for Education and Research in Information Assurance and Security (CERIAS)

and Department of Computer Science

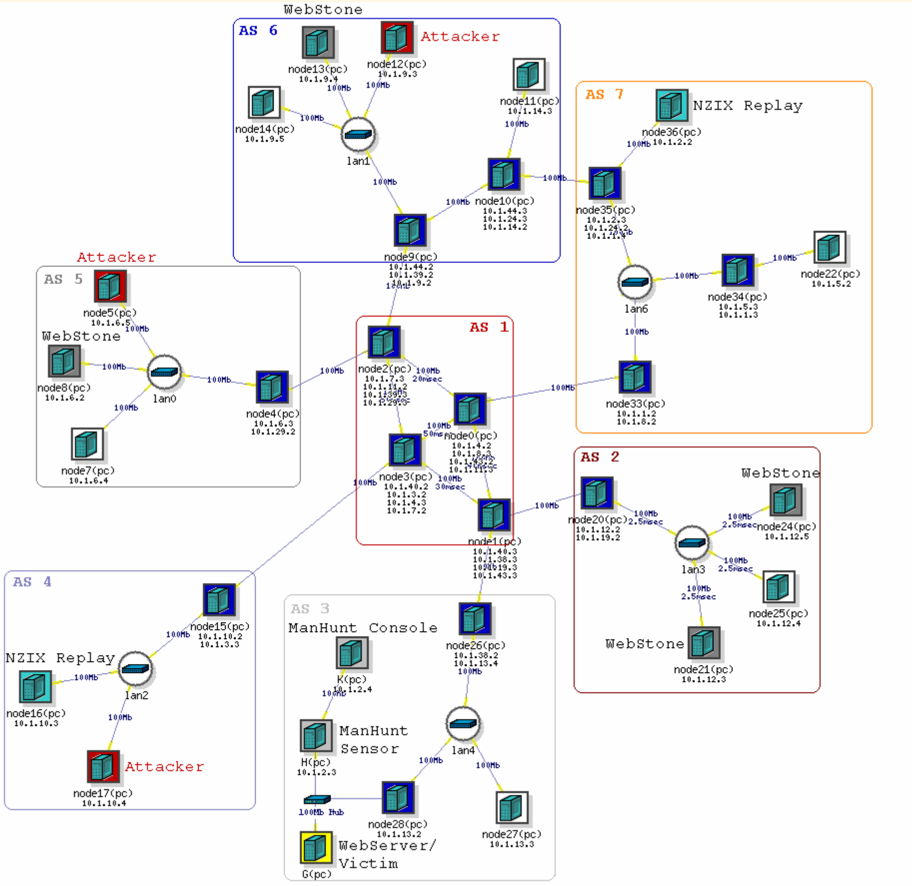
Purdue University

http://www.cs.purdue.edu/~fahmy/

May 10th, 2006

How to evaluate an idea?

Evaluate a new Internet service, e.g., CDN; a router architecture, e.g., QoS; a protocol, e.g., a new wireless CC protocol; or understanding the phenomenology of an attack, e.g., worms, BGP, or DDoS attacks



What approach should be taken to evaluate new ideas, or understand Internet behavior under various conditions?

Networking Research Process

Network Technology Development Cycle:

1. Come up with an idea and perhaps model it
2. Simulate the basic prototype
3. Use emulation/small testbed to build a release version
4. Deploy on the Internet or a private network

Use measurement results to build better simulation and emulation models.

PURDUE
UNIVERSITY

Examples

- Mathematical models, e.g., queuing models
- Simulation, e.g., ns-2, pdns, SSFNet, J-Sim, OPNET, GloMoSim
- Emulation testbeds (emulation tools: DummyNet, NIST-net, ModelNet, Wisconsin’s Click-based Path Emulation)
  - Emulab (www.emulab.net)
  - *DETER (www.deterlab.net) and EMIST (www.isi.edu/deter)*
  - WAIL (www.schooner.wail.wisc.edu)
  - VITELS (University of Bern+)
- Small (local-area) testbeds, e.g., *MAP*
- Wide-area testbeds, e.g., PlanetLab (www.planet-lab.org), MIT RON, VoIP testbed
- Production testbeds, e.g., Internet2, GEANT2, *e-stadium*
- GENI (www.geni.net)?

Wireless/Sensor Testbeds

- Typically exclusive access to a node/link
- Can be even harder to manage
- Examples
    - Kansei (OSU)
    - Orbit (Rutgers)
    - Emulab (Utah)
    - *E-stadium (Purdue)*
    - *MAP (Purdue)*

Simulation

- Layers
  - No real layers. Packets are treated as messages: ns-2, pdns
  - Real layers from layer 2 and up: GTNeTS, OPNET, OMNeT++
- Device models
  - General and simple (e.g., serv_delay = pkt_size / BW): ns-2, pdns, GTNeTS
  - Custom models *per device*: OPNET and OMNeT++
- Protocol Software base
  - Custom implementation: ns-2, OPNET, OMNeT++, pdns, GTNeTS
  - Relies on production code: Network Simulation Cradle add-on for ns-2, NCTUns

- Sam Jansen, Network Simulation Cradle http://www.wand.net.nz/~stj2/nsc/
- S. Wang et al., The Design and Implementation of the NCTUns 1.0 Network Simulator, Computer Networks 2003

Emulation

- Bridges simulation and the real world by providing network "clouds" to which physical components connect
- Can be used to shape links (DummyNet and Click) or emulate an entire network (ModelNet, EMPOWER, and VINT)
  - F. Baumgartner et al., Virtual routers: A Tool for Emulating IP Routers, LCN 2002/CCR 2003
  - L. Rizzo, DummyNet, http://info.iet.unipi.it/~luigi/ip_dummynet/
  - E. Kohler et al., The Click Modular Router, ACM TOCS 2000
  - A. Vahdat et al., Scalability and Accuracy in a Large-Scale Network Emulator, OSDI 2002
  - P. Zheng and L. Ni, EMPOWER: a Network Emulator for Wireline and Wireless Networks, INFOCOM 2003
  - K. Fall, Network Emulation in the Vint/NS Simulator, ISCC 1999
- Nodes can be virtualized on a single PC: vBET, Emulab NSE
  - X. Jiang and D. Xu, vBET: a VM-Based Emulation Testbed, MoMeTools 2003
  - B. White et al., An Integrated Experimental Environment for Distributed Systems and Networks, OSDI 2002

PURDUE
UNIVERSITY

Device Measurement

- Basic network device profiling metrics like: maximum throughput rate, packet loss, route setup, packet service time, and service recovery have been outlined in RFC 2544 and RFC 2889.
  - S. Bradner and J. McQuaid, Benchmarking Methodology for Network Interconnect Devices, RFC 2544, 1999
  - R. Mandeville and J. Perser, Benchmarking Methodology for LAN Switching Devices, RFC 2889, 2000
- Benchmarks in the above RFCs only deal with homogeneous traffic. Traffic representative of real networks induces different stresses.
  - J. Sommers and P. Barford, Self-Configuring Network Traffic Generation, SIGCOMM 2004
- Black box profiling has been done to measure OSPF calculations on Cisco routers.
  - A. Shaikh and A. Greenberg, Experience in Black-box OSPF Measurement, IMW 2001

PURDUE UNIVERSITY

Device Modeling

- Simulators and emulators can model a router device by using facilities like: variable delay, policies per packet, rate limiting, etc.
  - Most current tools do not do this and concentrate on general connectivity and output queuing models, in order to scale
- Simulators like OPNET/OMNeT++ have device specific models.
  - It is hard to manage a very large database of models.
  - A small change in the router's software can invalidate a previous model
  - Validation and accuracy
  - Complex models add large computational overhead
- Black box profiling.
  - Has been done in limited settings but no attempts to create a general model.
  - No policy derivation methods
  - *Is it possible to use profiling to create better simulation/emulation models?*

PURDUE
UNIVERSITY

Experimentation Methods

Method	Scalability	Fidelity	Ease of use
Simulation	Large scale	Problematic	Very easy, but how to test new boxes?
Emulation	Small-large	Emulated parts can be problematic	Requires expertise
Small testbeds	Small	High, but wirespools or delay emulators required	Hard to change; expensive
Wide-area testbeds	Medium+	High, but results are not reproducible if links are shared, and containment is problematic	Hard to change/manage; can raise liability issues

PURDUE UNIVERSITY

Key Challenges

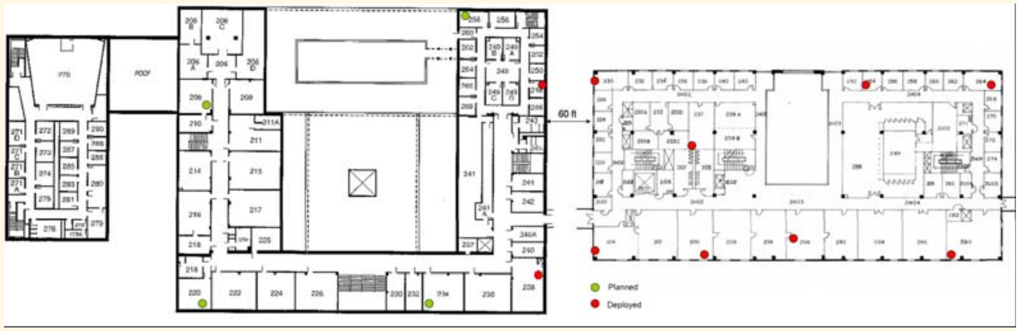
- Modeling:
  - Networks are composed of links and interconnecting devices which have various limitations and properties. If these limitations and properties are ignored in simulation and emulation models, then due to lack of *fidelity*, critical discrepancies between the tested and deployment behaviors can arise, e.g., DDoS, MediaPlayer
- Testbed design:
  - Isolated testbeds have no real users versus real users have expectations of privacy, availability, …
    - Reproducibility/containment versus benchmark *fidelity*
  - Shared use of the same node/link at the same time (e.g., through virtualization) in order to scale, versus artifacts and instrumentation difficulties
    - Understanding whether your results are due to the experiment versus platform is non-trivial when platform is complex
  - Diversity of devices on a testbed versus testbed manageability and security, and result reproducibility
  - Autonomy of different parts of the network versus manageability/security
- Scale down problem
- Researchers may reverse-engineer their solutions for the environment!

Case Studies

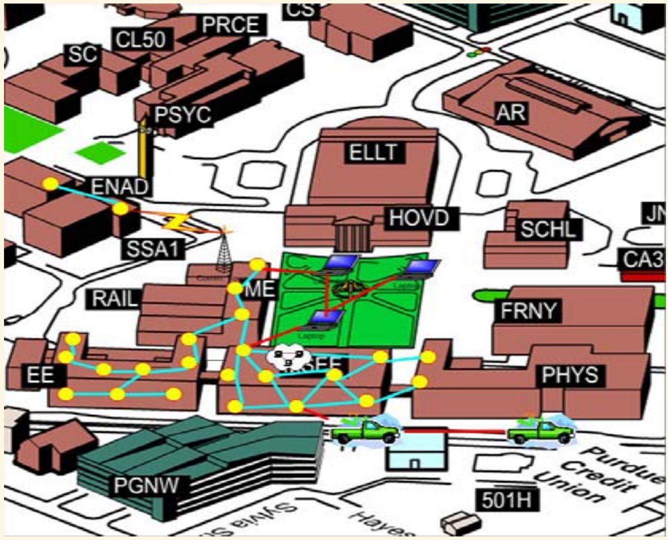
MAP: Wireless Mesh Network Testbed at Purdue (*Y. Charlie Hu*)

- Mesh Networks:
  - Wireless *multi-hop* networks
  - End devices, wireless routers, gateways
  - A packet reaches the gateway by going through multiple routers
  - Ad hoc, large-scale deployment
  - Promise to solve **the Last Mile** problem (home <-> local hub)
- MAP research activities
  - Networking: high-throughput via PHY/MAC/Routing/Transport
  - Systems: DHCP/Plug-n-Play/Monitoring/Security/Self-healing/QoS/…
  - Applications: Content sharing/ Streaming/Gaming/Location service/…

Phase 1 (current deployment)



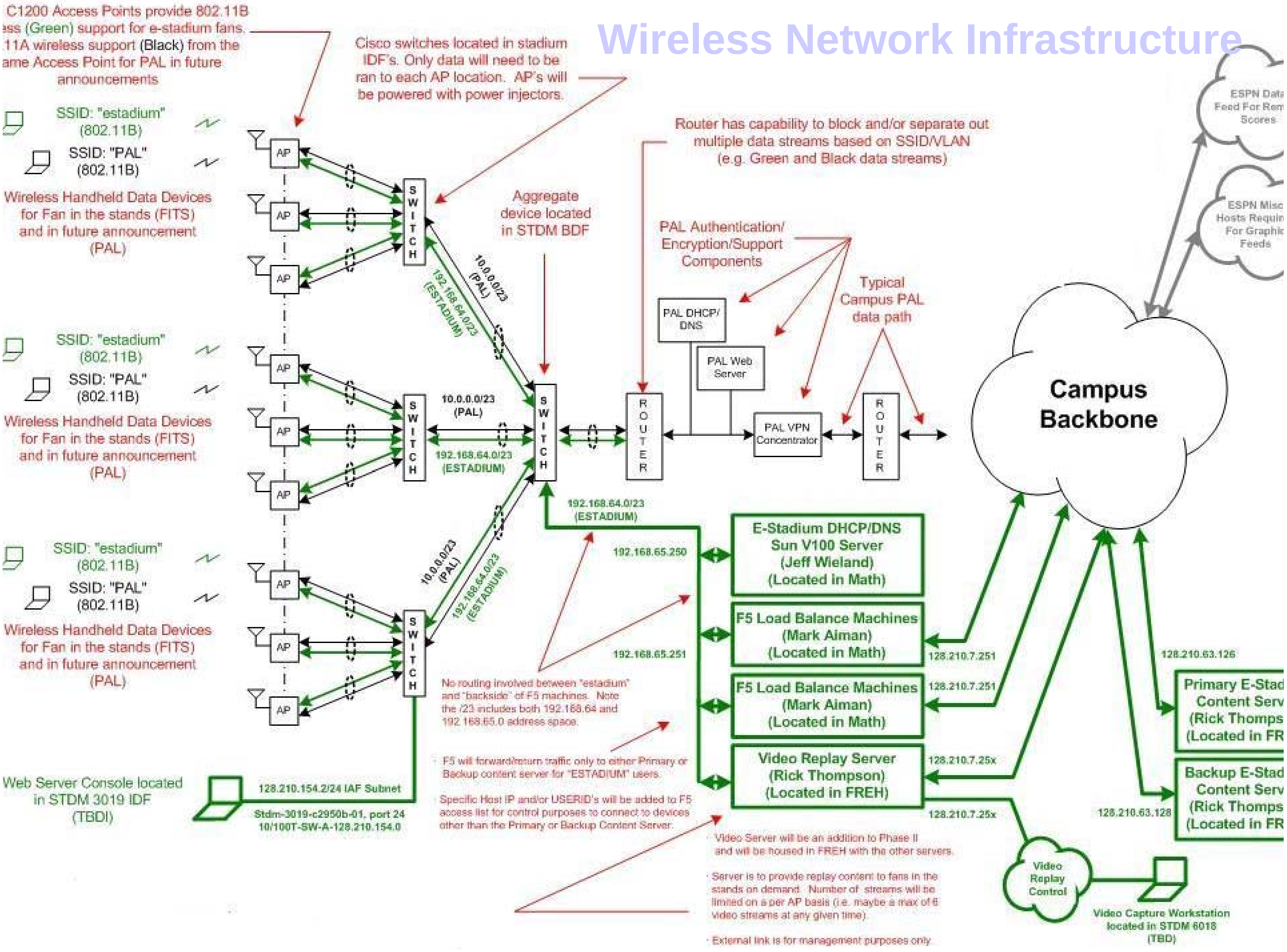
Phase 2 (planned expansion)



E-Stadium (E. Coyle)

- Create the most technologically advanced stadium in collegiate athletics
- Create a “Living Lab” for research + education in wireless networking
- Identify/solve problems in the *scalable* delivery of on-demand multimedia applications over wireless channels

Wireless Network Infrastructure



C1200 Access Points provide 802.11B ess (Green) support for e-stadium fans. 11A wireless support (Black) from the ame Access Point for PAL in future announcements

Cisco switches located in stadium IDF's. Only data will need to be ran to each AP location. AP's will be powered with power injectors.

Router has capability to block and/or separate out multiple data streams based on SSID/VLAN (e.g. Green and Black data streams)

Aggregate device located in STDM BDF

PAL Authentication/ Encryption/Support Components

Typical Campus PAL data path

No routing involved between "estadium" and "backside" of F5 machines. Note the /23 includes both 192.168.64 and 192.168.65.0 address space.

- F5 will forward/return traffic only to either Primary or Backup content server for "ESTADIUM" users.
- Specific Host IP and/or USERID's will be added to F5 access list for control purposes to connect to devices other than the Primary or Backup Content Server.

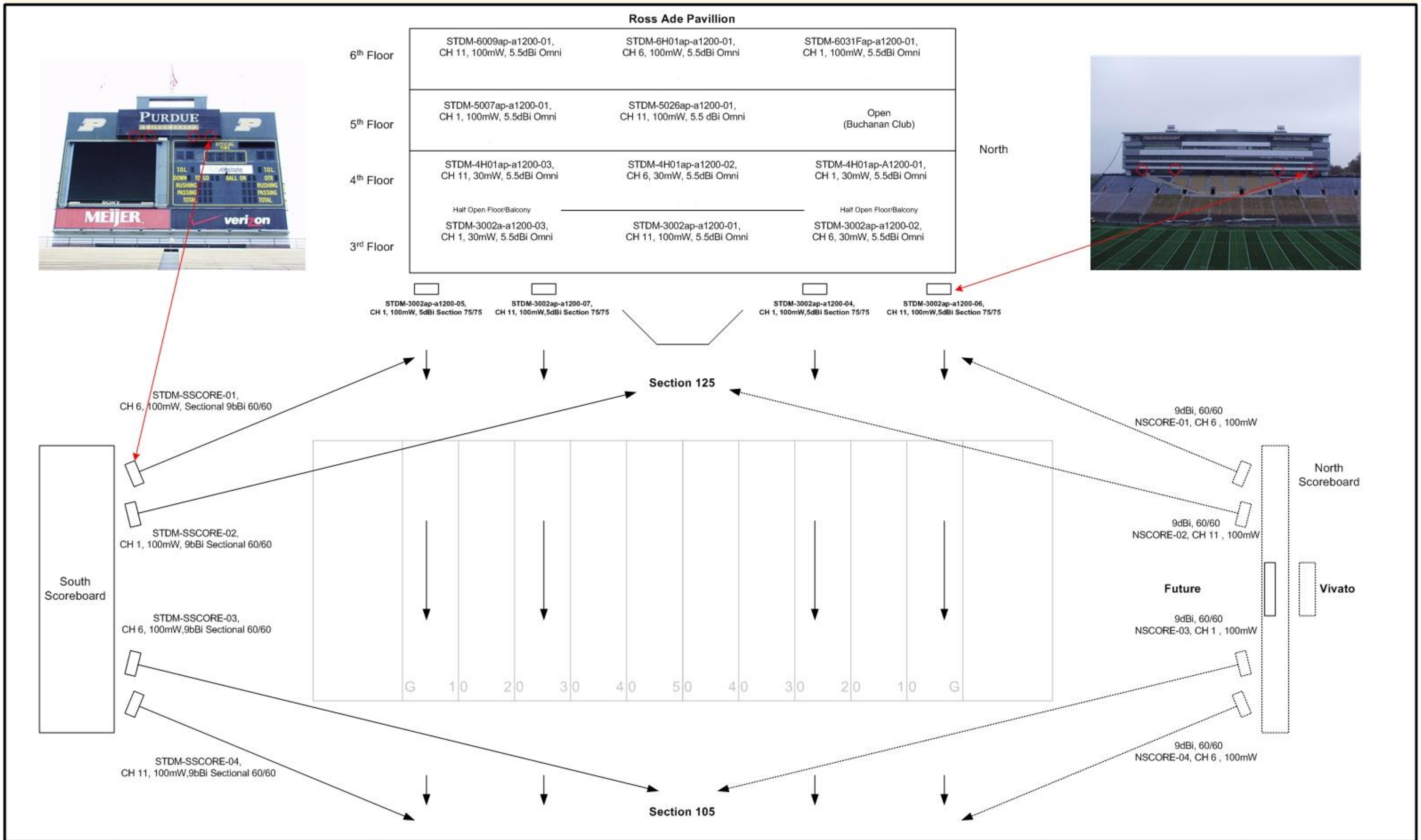
- Video Server will be an addition to Phase II and will be housed in FREH with the other servers.
- Server is to provide replay content to fans in the stands on demand. Number of streams will be limited on a per AP basis (i.e. maybe a max of 6 video streams at any given time).
- External link is for management purposes only.

Web Server Console located in STDM 3019 IDF (TBD)

Video Capture Workstation located in STDM 6018 (TBD)

SSID: "estadium"
(802.11B)
SSID: "PAL"
(802.11B)
Wireless Handheld Data Devices for Fan in the stands (FITS) and in future announcement (PAL)
SSID: "estadium"
(802.11B)
SSID: "PAL"
(802.11B)
Wireless Handheld Data Devices for Fan in the stands (FITS) and in future announcement (PAL)
SSID: "estadium"
(802.11B)
SSID: "PAL"
(802.11B)
Wireless Handheld Data Devices for Fan in the stands (FITS) and in future announcement (PAL)
AP
AP
AP
AP
AP
AP
AP
AP
AP
SWITCH
SWITCH
SWITCH
SWITCH
10.0.0.0/23 (PAL)
192.168.64.0/23 (ESTADIUM)
10.0.0.0/23 (PAL)
192.168.64.0/23 (ESTADIUM)
10.0.0.0/23 (PAL)
192.168.64.0/23 (ESTADIUM)
192.168.64.0/23 (ESTADIUM)
ROUTER
PAL DHCP/ DNS
PAL Web Server
PAL VPN Concentrator
ROUTER
Campus Backbone
ESPN Data Feed For Rem Scores
ESPN Misc Hosts Requir For Graphic Feeds
192.168.65.250
192.168.65.251
E-Stadium DHCP/DNS Sun V100 Server (Jeff Wieland) (Located in Math)
F5 Load Balance Machines (Mark Aiman) (Located in Math)
F5 Load Balance Machines (Mark Aiman) (Located in Math)
Video Replay Server (Rick Thompson) (Located in FREH)
128.210.7.251
128.210.7.251
128.210.7.25x
128.210.7.25x
128.210.63.126
128.210.63.128
Primary E-Stad Content Serv (Rick Thomps (Located in FR
Backup E-Stad Content Serv (Rick Thomps (Located in FR
Video Replay Control
128.210.154.2/24 IAF Subnet
Stdm-3019-c2950b-01, port 24
10/100T-SW-A-128.210.154.0

E-Stadium AP Layout



Ross Ade Pavillion

6th Floor	STDM-6009ap-a1200-01, CH 11, 100mW, 5.5dBi Omni	STDM-6H01ap-a1200-01, CH 6, 100mW, 5.5dBi Omni	STDM-6031Fap-a1200-01, CH 1, 100mW, 5.5dBi Omni
5th Floor	STDM-5007ap-a1200-01, CH 1, 100mW, 5.5dBi Omni	STDM-5026ap-a1200-01, CH 11, 100mW, 5.5 dBi Omni	Open (Buchanan Club)
4th Floor	STDM-4H01ap-a1200-03, CH 11, 30mW, 5.5dBi Omni	STDM-4H01ap-a1200-02, CH 6, 30mW, 5.5dBi Omni	STDM-4H01ap-A1200-01, CH 1, 30mW, 5.5dBi Omni
	Half Open Floor/Balcony		Half Open Floor/Balcony
3rd Floor	STDM-3002a-a1200-03, CH 1, 30mW, 5.5dBi Omni	STDM-3002ap-a1200-01, CH 11, 100mW, 5.5dBi Omni	STDM-3002ap-a1200-02, CH 6, 30mW, 5.5dBi Omni

North

STDM-3002ap-a1200-05,
CH 1, 100mW, 5dBi Section 75/75

STDM-3002ap-a1200-07,
CH 11, 100mW,5dBi Section 75/75

STDM-3002ap-a1200-04,
CH 1, 100mW,5dBi Section 75/75

STDM-3002ap-a1200-06,
CH 11, 100mW,5dBi Section 75/75

Section 125

STDM-SSCORE-01,
CH 6, 100mW, Sectional 9bBi 60/60

STDM-SSCORE-02,
CH 1, 100mW, 9bBi Sectional 60/60

South Scoreboard

STDM-SSCORE-03,
CH 6, 100mW,9bBi Sectional 60/60

STDM-SSCORE-04,
CH 11, 100mW,9bBi Sectional 60/60

9dBi, 60/60
NSCORE-01, CH 6 , 100mW

North Scoreboard

9dBi, 60/60
NSCORE-02, CH 11 , 100mW

Future

Vivato

9dBi, 60/60
NSCORE-03, CH 1 , 100mW

9dBi, 60/60
NSCORE-04, CH 6 , 100mW

Section 105

Information Technology at Purdue	ESTADIUM AP LAYOUT		Name: Doug Magers Job title: Network Engineer	PURDUE UNIVERSITY
Telecommunications	Date: 10/21/2004	Revision: 1.00 DRAFT Pg 1 of 1	Email: dmagers@purdue.edu	

Case Study: Security Experiments

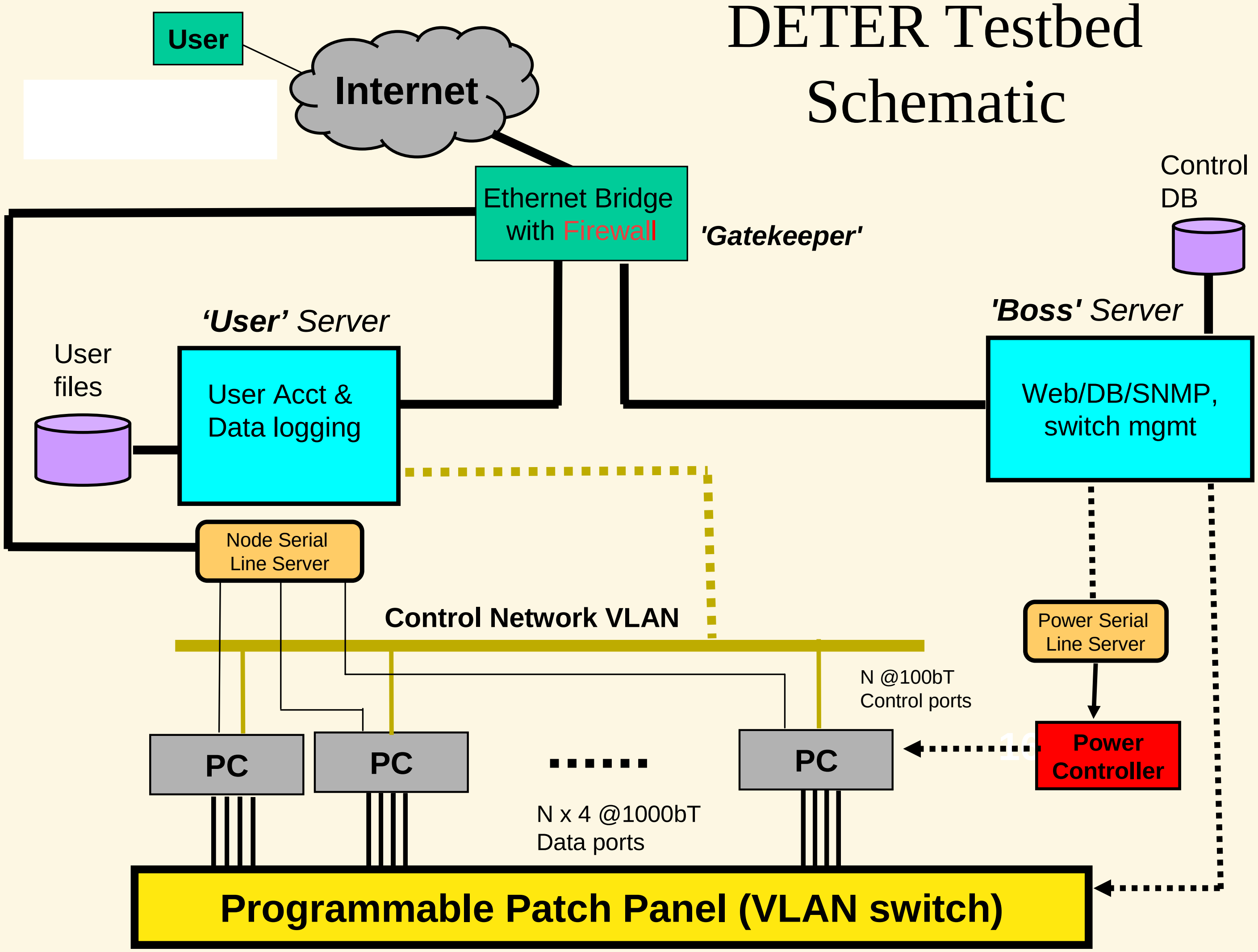
- Problem: Inadequate wide scale deployment of security technologies
  - Despite many years investment in network security research!
- One reason: Lack of experimental infrastructure
  - Testing and validation via simulations or in small to medium-*scale* private research labs
  - Missing objective and high *fidelity* test data, traffic, and metrics

➢ We still do not fully understand attacks and defenses in realistic settings of more than a few nodes!

DETER/EMIST

- A key goal of DETER/EMIST is to develop rigorous testing methodologies, tools, and benchmarks for important classes of Internet attacks and defenses.
  - It is crucial to understand the effectiveness of defense mechanisms on ***realistic*** networks (+stress tests).
  - ***Results obtained on testbeds can be used to develop more accurate analytical, simulation, and emulation models.***
  - Refs: Kohler and Floyd, Floyd and Paxson, … others.
- High *fidelity/scalability* is a key tradeoff
  - Simulators cannot execute *real applications/system software*, and only approximate various *appliances, e.g., IDSs*.
  - Emulation provides a convenient way to use ***real*** appliances and systems, though it is constrained by the *number of nodes, types of appliances,* and *difficulty in configuration/management/reproducibility*.
  - ***DETER is based on Emulab***

DETER Testbed Schematic



Source: DETER USC-ISI team, based on Utah Emulab

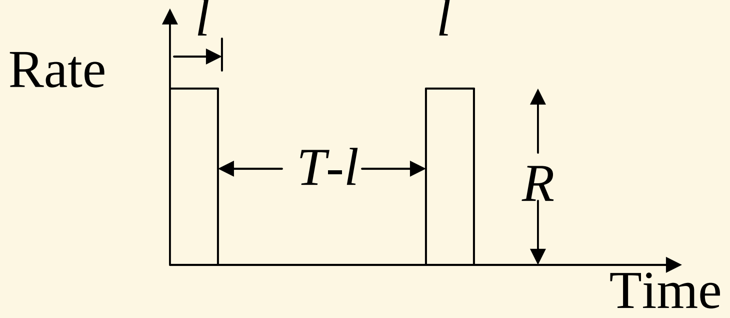
Tools

- *Large scale* experiments on an emulation testbed require *(i) topology generation, (ii) extensive router configuration, (iii) automated node control with synchronization, and (iv) support for sensitivity analysis.*
- Hence, it is important to create an infrastructure for fast experiment creation and automation, including complex BGP/OSPF scenarios.
- http://www.cs.purdue.edu/~fahmy/software/emist/ contains many control, measurement, logging, and visualization tools.

PURDUE UNIVERSITY

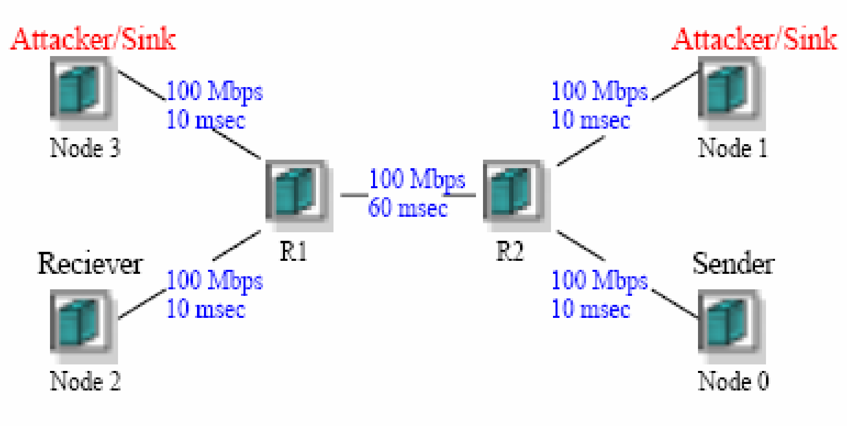
TCP-Targeted Attacks

- Why? Easy to *launch, stealthy,* and potentially damaging attack
  - **A. Kuzmanovic and E. W. Knightly. Low-rate targeted denial of service attacks. SIGCOMM 2003.**
  - **H. Sun et al. Defending against low-rate TCP attacks: Dynamic detection and protection. ICNP 2004.**
  - **M. Guirguis et al. Exploiting the transients of adaptation for RoQ attacks on Internet resources. ICNP 2004.**
- Studied *only* via simulation and limited experiments
- Tricky as it strongly relies on timing (phase effects)
- Vary: Attacker, burst length *l,* sleep period *T-l, pkt size, RTT, bfr size*
- Objective:
  - Understand attack effectiveness (damage versus effort)
  - *Qualitatively* compare emulation to simulation to analysis



Simplest Experimental Scenario

- Original TCP-targeted attacks are tuned to RTO frequency for near zero throughput
- Can exploit Additive Increase Multiplicative Decrease congestion avoidance of TCP *without* tuning period to RTO, and hence throttle TCP's throughput at any predetermined level
- Simple dumbbell topology with single file transfer flow is easiest to interpret and is the most demanding for attacker



PURDUE
UNIVERSITY

Throughput Degradation

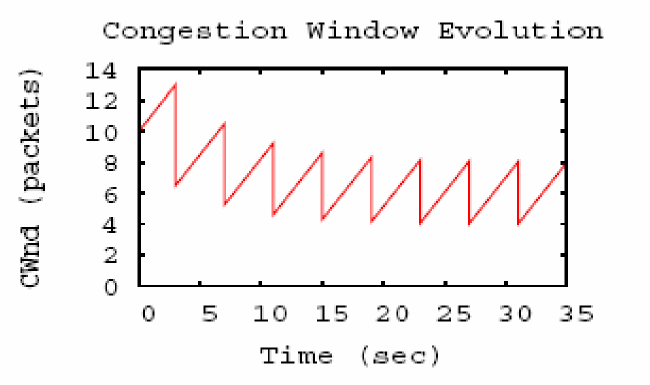
- Loss occurs during each pulse.
- Connection does not RTO.
- There is no packet loss during attack sleep periods.

$$W_{i+1} = \frac{W_i}{2} + \alpha$$

$$W_3 = \frac{\frac{\frac{W_N}{2}+\alpha}{2} + \alpha}{2} + \alpha$$

$$W_S = \lim_{i\to\infty}\left[\left(2^{-i}W_N + \alpha\left(\sum_{j=0}^{i-1} 2^{-j}\right)\right) = \alpha\left(\sum_{j=0}^{i-1} 2^{-j}\right)\right] = 2\alpha$$

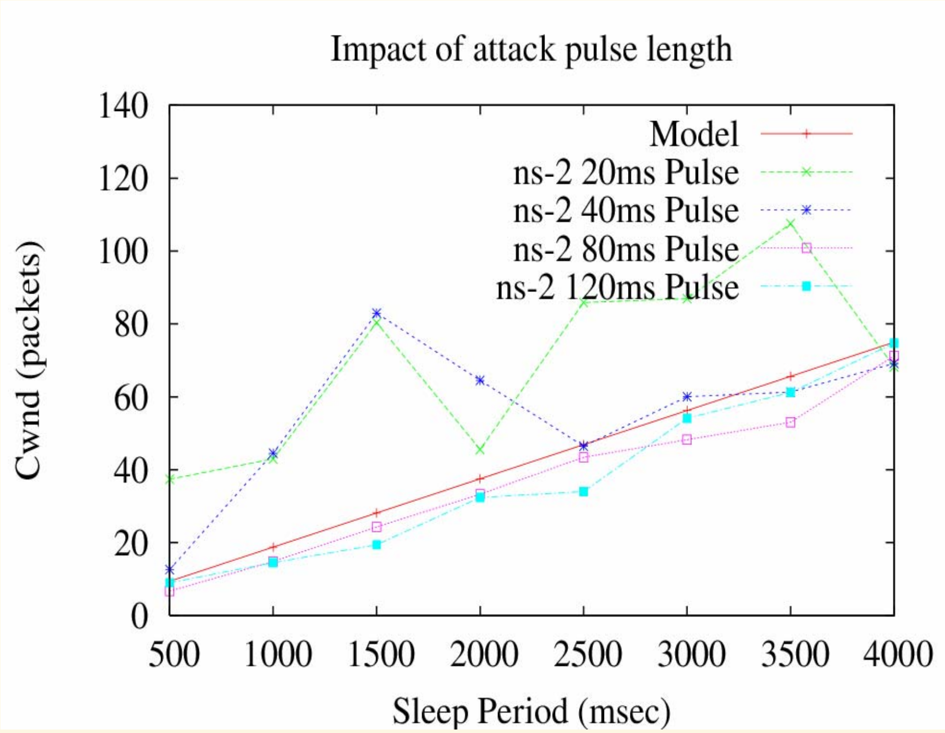
$$W_{avg} = \frac{3t}{4rtt}$$

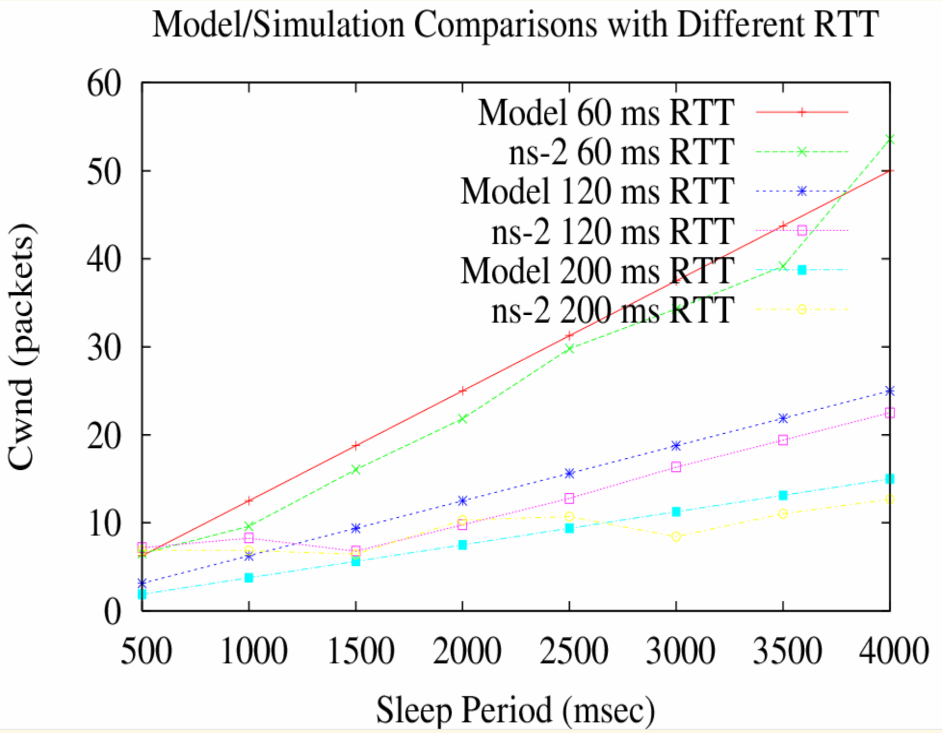


α is the Cwnd growth during a sleep period

t time between two loss events

Analysis vs. Simulation

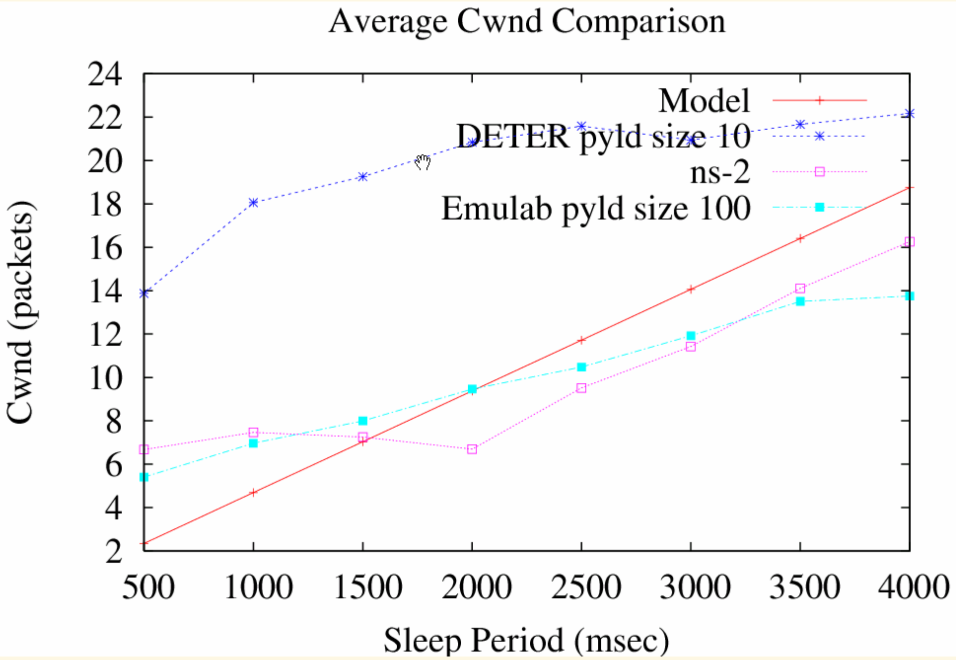


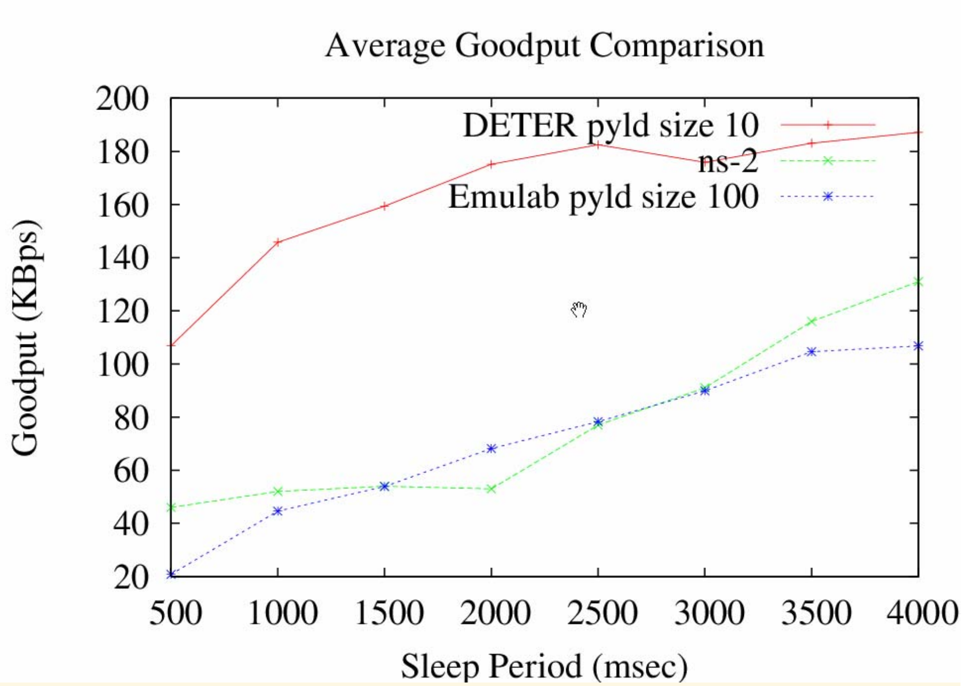


- Simulation results are closest to the analysis when the attack pulse length is equal to the flow RTT.
- Non-monotonic increase amplified by ***phase effects***.
- Adding randomization helps.

PURDUE UNIVERSITY

Forward Direction





- Analysis matches ns-2 results when attack pulse length is greater or equal to TCP flow RTT and when buffer sizes are not too large
- DETER is not as affected by the attack: Why?
- **Experiments with WAIL show that *PC routers* outperform Cisco 3640 dep. on settings (consistent with results reported by several companies).**
- Such differences are important as they allow us to identify *real vulnerabilities and fundamental limits,* e.g., NAT boxes, combination of capabilities.
- The Internet is an evolving, ***heterogeneous*** entity with implementation errors and resource constraints, and not an approximation in a simulator or a uniform emulator

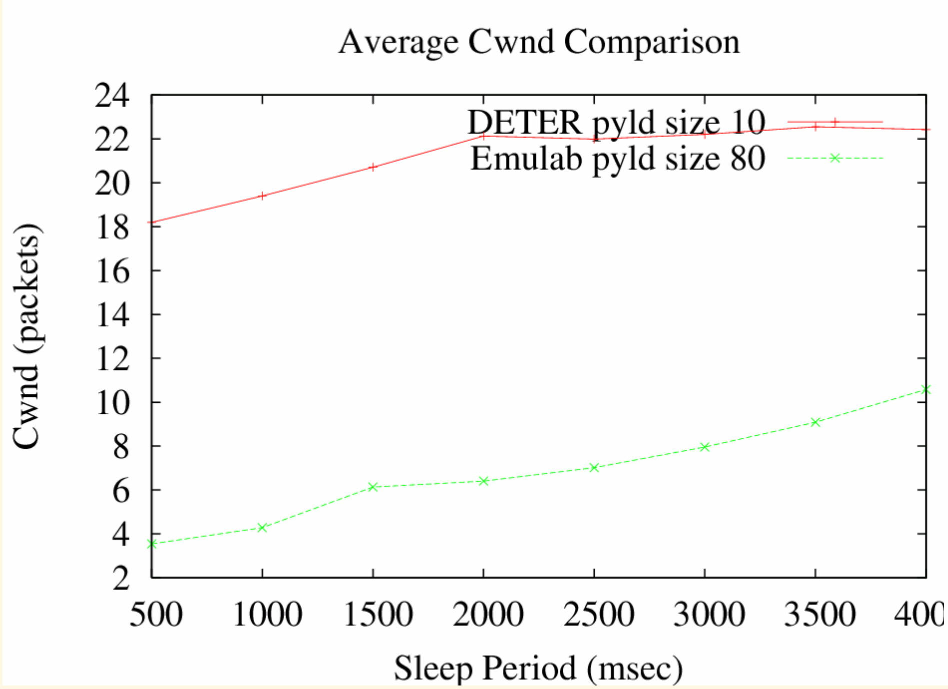
Receive Livelock

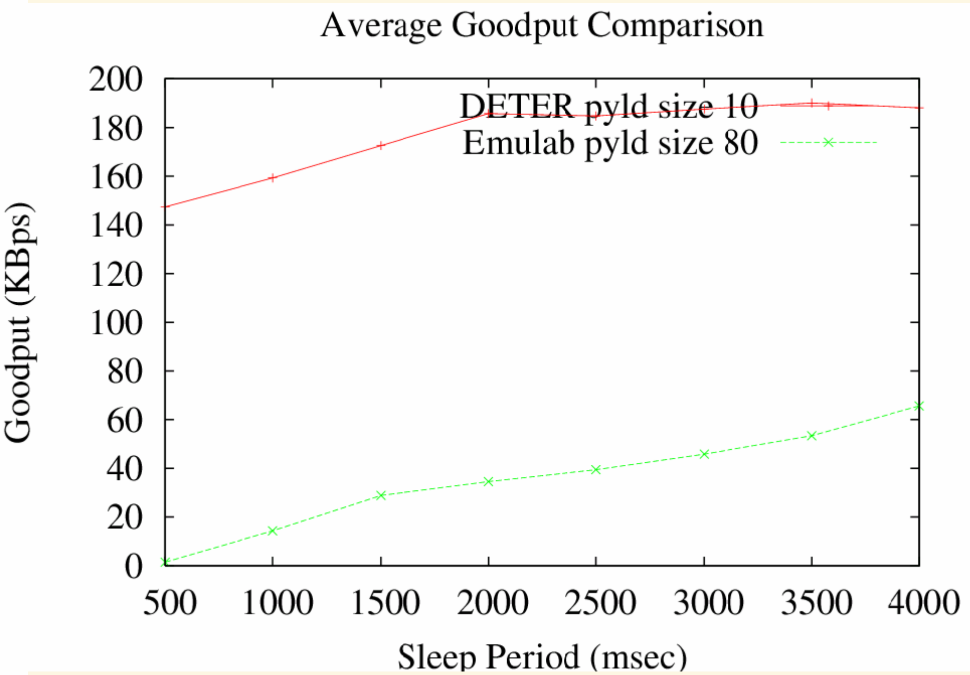
- Schemes that receive packets by invoking interrupts suffer from:
  - High CPU utilization
  - Reduced forwarding rate
  - Process starvation

- Polling solves the above problems by:
  - Using software interrupts and a kernel thread reduces interrupt overhead by batching the receive signals
  - Batch limits govern the time the CPU spends in kernel mode processing the packets

- J. Mogul et al., Eliminating Receive Livelock in an Interrupt-driven Kernel, ACM Transactions on Computer Systems, 1997
- P. Druschel et al., Experiences with a High-speed Network Adaptor: A Software Perspective, SIGCOMM 1994
- Kohler et al., The Click Modular Router, ACM TOCS 2000

Reverse Direction





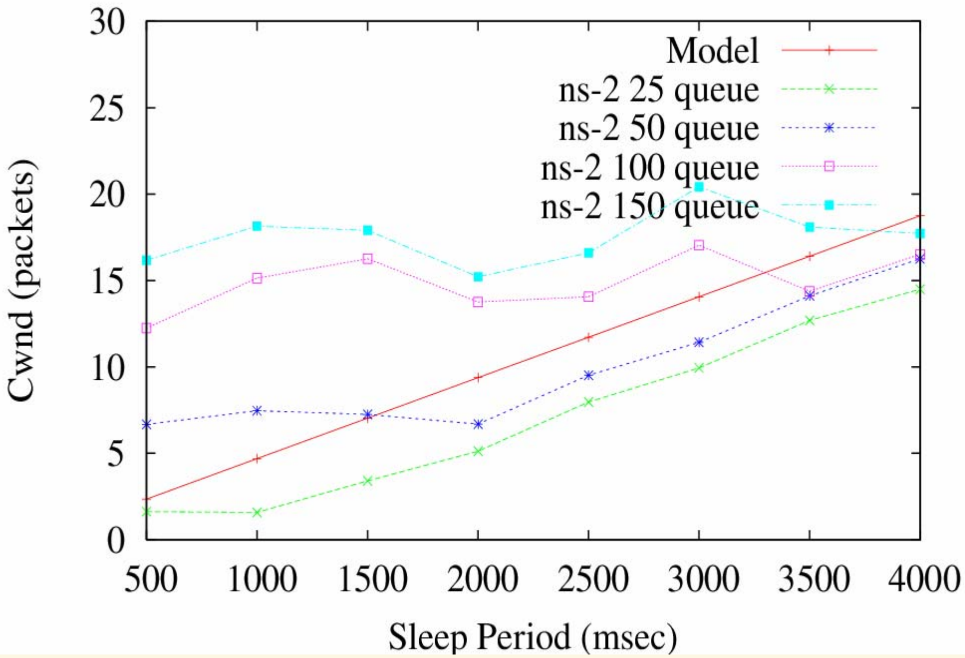
- Since ns-2 does not model CPU/bus/devices, and opposing flows do not interfere at a router with output buffering, data for ns-2 is not shown for reverse direction (Cwnd has no cuts)
- Cisco 3640s or NATs would behave differently

Router Nodes

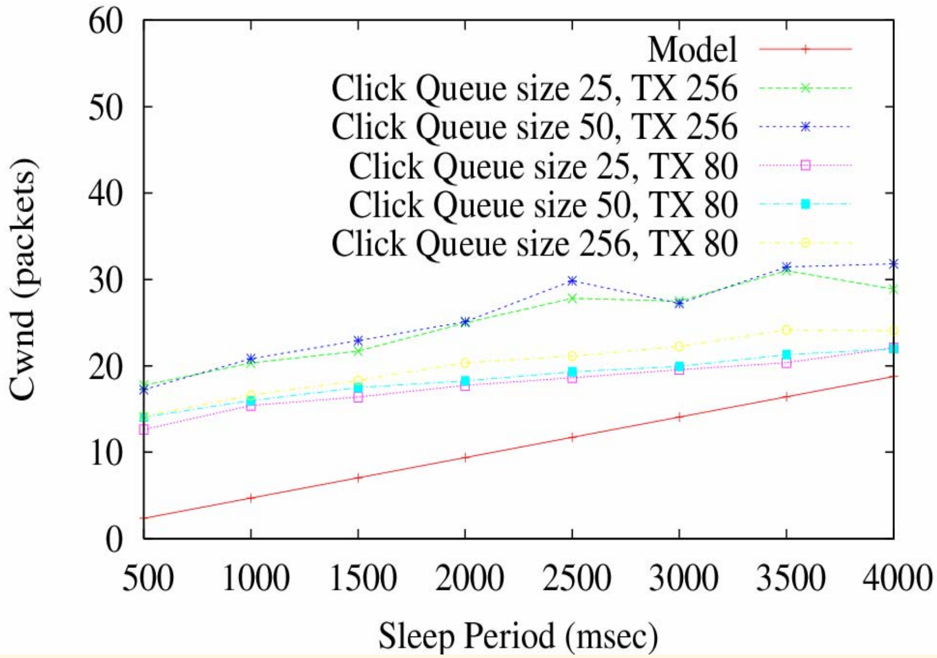
- To avoid slowdown in the Linux kernel, the machine can be configured to run SMP enabled Click modular router with polling drivers.
  - Polling reduces overhead by reducing interrupts [Mogul97, Druschel94, Kohler2000].
  - Bypassing the Linux protocol stack speeds up packet processing.
  - It is important to carefully select and configure delay nodes to ensure no drops!
- It is important to configure *network device buffers* as well, since default values are unreliable.

Results with Click

Impact of router queue size



Impact of Click and Driver Queue Size



- The results indicate that device buffer size variation has a higher impact on the final results than Click buffers.
  - *It is important to understand device drivers so that accurate comparisons with real routers can be made.*
- *Differences between different routers need to be modeled!*

PURDUE
UNIVERSITY

Results with Cisco 3640

- We had to use TCP packets instead of UDP as the router’s *policy* gives preference to TCP over UDP packets.
- The attack rate was limited to *Maximum Loss Free Receive Rate (MLFR)* to avoid significant input queue packet loss.
- Contrary to previous results *larger packets at lower rate* caused more damage.
- Cisco 7206VXR has significantly different behavior.

Summary of DDoS Results

- Results are *highly* sensitive to attack and scenario parameters – need more control in resource assignment on shared testbeds and more care with hw/sw upgrades
- Differences between DETER, WAIL, and Emulab testbed results *with similar configurations and identical scripts* are attributed to differences in the underlying hardware and system software, especially NICs/device drivers, and buses.
- Click experiments demonstrate the importance of device driver *settings*.
- *Can we use Click and device driver options as well as relative node capabilities to quickly and approximately emulate DDoS scenarios with popular routers on the Internet today, e.g., Cisco 36xx, 7xxx, GSR 12xxx, Junipers, … etc?*
- *Can simulators model such differences in a scalable manner?*