

How to Get a Systems PhD in Three Years

Diomidis Spinellis
<http://www.dmst.aueb.gr/dds>

Presentation outline

- Computing Systems Research
- Thesis Outline
- Why Implement
- How to Implement
- Obtaining Feedback
- References
- And Life

What is a Systems PhD?

- Information Systems context
 - Society, community, organization
- Computing Systems context
 - Software, hardware



How to select a problem

- Change way people do Computer Science & Engineering
- Solve a problem people care about
- Short project
 - Always takes longer than expected
 - Limited time (1.5 yr)
- Strive for multidisciplinary work
- Match our local environment (+/-)
- Should keep you excited for 3-4 years



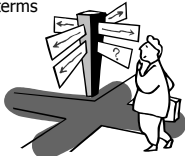
The Future of Systems Research

- Availability
 - Of appliances and services
 - Maintainability
 - Prevent failure, make recovery easy
 - Scalability
 - Satisfy user demand
 - Follow technology evolution
- (Hennessy 1999)



Thesis Outline

1. Introduction **Yr3**
 - Written last, thesis banner ad
2. Evaluation framework and related work **Yr1**
 - Convince others you know your field
3. Theoretical work, approach description **Yr3**
 - Description in the most general terms
4. Plan of attack **Yr3**
 - Future tense
5. Work done **Yr2**
6. Evaluation **Yr1-3**
 - Use evaluation framework
7. Future work **Yr3**
8. Conclusions **Yr3**



Why implement?

- Proof of concept
 - Designs and ideas are cheap, implementations are expensive
 - (particularly those based on unsound designs)
- Explore alternatives
- Performance measurements
- Implementation metrics
- Proof by community acceptance
- Obtaining Feedback



Proof of Concept

- Virtual Services: A New Abstractions for Server Consolidation (Reumann et al, 2000)
- ASPs multiplex server resources among multiple clients
- Applications may interfere
- Establish Virtual Services (VSs)
- VS Gate: modified Linux kernel to classify system calls according to VS membership
- "We demonstrated that VSs are an effective, application-oriented transparent resource management abstraction when subservices are shared across business clients"



Exploring alternatives

- Performing Replacement in Modem Pools (Smaragdakis and Wilson, 2000)
- Modems are a time limited resource
- Explore between
 - LRU
 - CIRG (conditional inter-reference gap)
 - Match user patterns
 - Random
- CIRG was found to be the best predictor of future idle times



Performance Measurements

- Towards Availability Benchmarks: A Case Study of Software RAID Systems (Brown and Patterson, 2000)
- Task: availability of RAID implementations
- Waiting for faults impractical
- Create emulated disk with fault workload
- Quantitative evaluation
 - Linux (slow, but no performance drop)
 - Windows 2000 (in the middle)
 - Solaris (fast, impacts performance)
- Qualitative evaluation (restoring service)



Metrics

- Sandboxing Applications (Prevelakis and Spinellis 2001)
- Many applications (e.g. web browsers) are insecure
- Secure operation is inflexible (e.g. disabled JS)
- Use an OS-supported sandbox for arbitrary applications
- Tool to create sandbox rules
- Therefore sandbox applications
- "Allowing the FMAC tool to learn from the trial runs of the application, reduces the initial work required for constructing the sandbox"



Proof by community acceptance

- SWILL: A Simple Embedded Web Server Library (Lampoudi and Beazley 2002)
- Adding a web interface is time consuming
- Especially for legacy applications
- Many applications do not fit the mold of an internet application
- With 3 SWILL calls your app serves web pages
- Used for
 - Scientific simulation monitoring
 - Operating systems student project
 - Compiler parse tree browsing



Feedback, Ideas in Practice



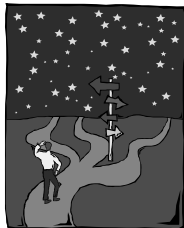
- "Our weather prediction system is up and running and no one has complained about its occasional inaccurate forecasts"

vs

- "Every time we fail to forecast rain, the users hang their wet shirts over their PCs to dry"
- (People use and depend on the system)

Implementation Alternatives

- Modify existing system
 - + Avoid duplication of work
 - + Ride on the bandwagon
 - Difficult to distinguish your contribution
- Prototype
 - + Quickly obtain results
 - Results may be unrealistic
 - Impossible to disseminate
- Build from scratch
 - + Best way forward
 - A lot more work
 - Difficult to disseminate



How to Implement (1)

- KISS
- Careful choice of innovation points
- Compatibility in everything else
- Complexity problems
 - Longer design + implementation
 - Can become outdated
- Use intuition to ask questions not answer them



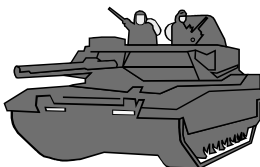
How to Implement (2)

- Experiment early and often
 - Against specification
 - Design alternatives
 - Implementation alternatives
- Keep checkpoints (CVS/RCS)
 - Lessens inhibitions to change
- Document
 - Workbook (vs)
 - Comments
 - Alternatives, progress, failures



Use Appropriate Tools

- Scripting Language
 - Perl, Ruby, Python, sh
- Compiled Language
 - C, C++, Java
- GUI front end builder
 - .net, swing, tcl/tk, QT, SWILL
- Open source OS
 - FreeBSD, OpenBSD, Linux



The Researcher's Toolchest

- Text editor, programming environment
- Generate summary reports out of logs and events
- Draw graphs and diagrams (automatically)
 - GNUPlot, graphviz
- Format text for the thesis and the web



Obtaining Feedback

- Networking
- Distribute implementation
- Web page
- Faculty, students, guests
- Seminars
- Publications



Evaluating your Work

- How does the system scale based on our framework?
- What should the reader learn from our system?
- How generally applicable are the lessons?
- Retest your initial assumptions against the built system
- How sensitive are the results against perturbations in the assumptions?



Writing Style

- How many have read Strunk & White?
- Is the writing clear and concise?
- Are words spelled and used correctly?
- Are the sentences complete and grammatically correct?
- Are ambiguity, slang, and cuteness avoided?



Spot Blind Spots in References

- No old papers
 - Possibly rediscovery of the wheel
- No new papers
 - Repeat recently published work
- Unrefereed material
 - Narrowness, parochialism



Reference Logistics

- Are your references in a database?
 - In your work
 - In your drawers / boxes
 - In the database
 - On disk
- Are the reference details complete and correct?
 - <http://www.cmsd.aueb.gr/dd/res/cite/index.htm>



Getting (and Keeping) a Life

- When did you last read (fnish) a non-work related book?
- Any RSI problems?
- Are you exercising?
 - (apart from your fingers online)
- Eating a healthy diet?
 - (apart from Pringles™)
- Attended a cultural event?
 - (not a football game)
- Friends / activities outside our group?
- Are we having fun / leading a meaningful life?



Self Evaluation



- Thesis pages written
- Papers read
 - Oldest
 - Newest
 - % online transient vs archival refereed
- When did you last browse our library's periodical collection?
- When was your last coding or implementation task?



References



- David A. Patterson. How to have a bad career in research/academia. Available online <http://www.cs.utah.edu/lepreau/osdi94/keynote/abstract.html>, November 1994. First Symposium on Operating Systems Design and Implementation (OSDI '94) keynote address, November 14-17, 1994 Monterey, California, USA.
- Roy Levin and David D. Redell. An evaluation of the ninth SOSP submissions or how (and how not) to write a good systems paper. *Operating Systems Review*, 17(3):35-40, July 1983. Also available online <http://gatekeeper.dec.com/pub/DEC/SRC/publications/levin/SOSPho wto.html>
- Phil Agre. Networking on the network: A guide to professional skills for phd students. Available online <http://dlis.gseis.ucla.edu/people/pagre/network.html>. Current January 2003, August 2002.
- Aviel William Strunk Jr. and E. B. White. The Elements of Style. Macmillan Publishing Co., 1979.

More References



- Chad Perry. A structured approach to presenting PhD theses: Notes for candidates and their supervisors. In *ANZ Doctoral Consortium*, Toronto, Canada, February 1994. University of Sydney.
- Steven Alter and Alan R. Dennis. Selecting research topics: Personal experiences and speculations for the future. *Communications of the Association for Information Systems*, 8:314-329, 2002.
- John Mingers. The long and winding road: Getting papers published in top journals. *Communications of the Association for Information Systems*, 8:330-339, 2002.
- James C. Wetherbe. Research center models for attracting corporate funding. *Communications of AIS*, 7(7), August 2001.
- A.R. Dennis and J.S. Valacich. Conducting research in information systems. *Communications of AIS*, 7(5), 2001.
