# Summary of Third Cross-Layer Reliability Meeting (Oct 29-30, 2009, Austin, TX)

## Goals

This final study meeting had two goals: understanding the constituency groups that had not presented at previous meetings, and crafting a plan for how the results of the study would be written up and presented to the CCC and funding agencies. Presentations from the life-critical systems and infrastructure working groups outlined the key issues facing their communities and some overlaps with other communities. Program managers from the NRL, NSF and DARPA attended the meeting, and provided feedback on how the study's results could be made most useful to them. A number of participants suggested that the study group propose a multi-agency program to fund cross-layer resilience research, and much of the later discussion focused on ways to pursue this suggestion.

## Tuning up Story

The workshop started in the same manner as the previous two workshops by telling the cross-layer visioning story (slides). Presenting the 10-20 slide story allows us to provide a basis for first time attendees and to refine the story to be told to funding agencies. As always, this presentation starts the discussion on reliability challenges and cross-layer reliability approaches. There were a number of suggestions that the participants provided:

- They pointed out that we should more crisply define the goals of the research. The suggestion was to show how the errors in logic could be handled by higher layers.
- They pointed out that NSF has a strong education focus.
- They pointed out that they needed more context on the fault rate so that lay people could understand what the fault rate means. For example, does it mean that one will need to reboot their computer every hour or that one will replace their processor every 10 minutes or that only 1% of hardware devices will yield?
- They pointed out that the mission impacts should include economic factors, as well as energy. In this part of the discussion the ability for the U.S. to compete in the global economy came up several times.
- They pointed out that all of this work should be communicated as a revolution rather than a revision.

## Framing for public: Immuno-Logic

One of the breakout sessions focused on ways to sell this type of project to the lay people. This discussion focused on two different objectives: finding slogans that would connect with the non-technical populace and how to sell the story. The clear winning concepts on slogans was "Immuno-Logic." Everyone felt that the immune system analogy was a good concept for cross-layer reliability, as both the human immune system and cross-layer reliability are multi-layer defense systems. Furthermore, lay people have a basic understanding of the immune system and understand how detrimental diseases that directly affect the functionality of the immune system, such as Leukemia and AIDS, are. Most people also have an understanding that the human immune system is innate and adaptive, which are two properties that we want computing systems to have. In both cases, the first line of defense is at the physical layer (devices reliability for circuits and physical barriers that keep pathogens out for the immune system) with additional, usually higher layer mechanisms, addressing the attacks that get past the physical defenses.

Being able to effectively sell the cross-layer reliability story to funding agencies and congress is necessary for further research progress on this topic. The discussion here focused on methods of protecting US-based technical companies/jobs and protecting us. The technology industry for several years felt pressure to outsource technical work to China and India. Many US-based companies outsource technical work to these countries to remain competitive in the global technology economy. The effects of this shift can be seen in both the increase in off-shore fabrication of silicon devices and the increase of off-shore electronics companies. Many participants stated that increasing the reliability of US-designed computing systems would help create value in US-built computing systems, increase the competitiveness of US-based companies, and increase jobs in the US technology market.

There is also a very compelling story to be told in how our computing systems protect us. As stated in later sections, the cost of reliability failures in automobiles, medical-implantable devices, and the energy infrastructure can be quite high. Reliability failures in these arenas can be expensive both in terms of human lives lost, but economically, too. Fairly trivial reliability failures in medical-implantable devices can lead to surgery to have the device explanted and replaced with a new device. In 2003 a cascading failure in the OH power infrastructure ended up affecting the entire northeastern US and Canada, which left 55 million people without power, played a role in 11 fatalities, and cost an estimated $6B [http://en.wikipedia.org/wiki/Northeast_Blackout_of_2003, http://www.scientificamerican.com/article.cfm?id=2003-blackout-five-years-later]. For our society to continue to embrace automation in banking, medical devices, automobiles, and infrastructure, the average, non-technical person needs to feel comfortable that the automation increases and not decreases their overall safety. Finally, we rely heavily on computational support for persistent surveillance for treaty monitoring of both the comprehensive test ban treaty and environmental treaties, as well as

warfighter support for the wars in Iraq and Afghanistan. Reliability failures in these arenas can cause fatalities in the battlefield, lead to bad policy decisions, and create confusion in the geo-political arena [http://en.wikipedia.org/wiki/South_Atlantic_Flash].

Following the Immuno-Logic theme, we discussed the value of a *public health* system for electronic systems. Ideally, we have designed each hardware system with adequate adaptive responses to deal with any reliability problems that may arise. However, we may encounter new reliability effects in the field that challenge the Immuno-Logic response of a single system. For these cases, some centralized data collection and dissemination could further protect systems in the same way that the public health system provides data collection and early warning of epidemics. We could even imagine developing inoculations to upgrade a system so it is better able to combat newly manifested reliability challenges.

# Government/Strategy

Discussions with NSF suggested we should engage the Engineering division as well as CISE in this topic and that we would get a better reception if we could get SRC (Semiconductor Research Corporation) involved as a partner. The topic should be interesting to Engineering, and it would be valuable to garner their support as well within NSF. Getting engagement from both Engineeering and CISE would help demonstrate the cross-cutting nature of the work as well as the broad impact. SRC involvement would demonstrate an industry commitment which would help communicate the importance and relevance of this issue to NSF. SRC's involvement may also help with some aspects of the program that NSF could not do as well on their own.

This naturally led to discussions on how to engage SRC. SRC is very responsive to the interests of its industrial members and always interested in expanding the industrial members involved. IBM, Intel, and Freescale all had participants at this meeting and are all SRC members. There was agreement to gather quotes from key principals at these companies to help make the case to the SRC for the interest of this topic to their industrial members. There was also discussion of these companies directing some of their SRC funding toward this theme; this would help provide some seed money for initiating funding of efforts in this area. CISCO is a potential new member for SRC, has participated in this study, and has interests in these issues.

# Education

The participants also had an open discussion regarding education, as many stated that resilience is not being taught currently in the EE and CS curricula. One participant pointed out that system reliability is taught as a discipline to

mechanical and civil engineers, so there is a precedence of teaching these types of ideas to undergraduate engineers. Many people also pointed out that we needed to start thinking about how to teach system reliability to computer engineers, including how to work on the K-12 pipeline (e.g. robotics, cubesat projects, and competitions). Several people also felt that there could be competitions tied to conferences, such as the branch predictor competition that was tied to MICRO 37 and 39 [http://www.jilp.org/cbp/, http://cava.cs.utsa.edu/camino/cbp2/]. For continued discussion on this topic, we have added a new wiki page to the relxlayer website to brainstorm educational opportunities [http://www.relxlayer.org/Education].

# Research Organization

At this meeting, we introduced a discussion on research organization (slides) that NSF brought up when they met with Nick, Heather, and André in September. Because the work that is needed to be done crosses the entire hierarchy of the computing system, the research needs to be cross-cutting work, demanding collaboration across disciplines and teams. This might necessitate big teams or centers to make progress and goes against funding models that focus on single-domain projects. Serialization of the research is also not possible, as getting an accurate model of device effects depends on a working architecture/software implementation in the technology. Two areas were discussed as possible near-term funding opportunities -- standard platforms/models and benchmarking -- as progress in these areas will provide the basis for later research.

There was suggestion that there are many similarities between our cross-layer reliability structure and the relatively modern power management infrastructure. Modern power management provides monitoring and control hooks to higher levels of software. Perhaps there is an opportunity to leverage some of the power management infrastructure in bootstrapping our tools and platforms? Perhaps there are lessons from the research and adoption of those techniques relevant to our problem?

# Life Critical

The life critical group briefed the workshop for the first time at this meeting. This group had two brief ins -- one from automotive (slides) and one from medically-implanted devices (slides). Both of these groups are regulated by the IEC 61508 standard, which is an international standard for "safety-related devices" [http://en.wikipedia.org/wiki/IEC_61508]. The automotive industry is also using a draft standard ISO 26262, which is an attempt to clarify IEC 61508 as it applies to automotive. Because of the safety concerns, these industries deliberately forgo advanced technology until the larger commercial industry determines how scaling affects the reliability of the technology. They also pointed out that they would benefit from more publicly-available operational data from existing

technologies. Currently, automotive technology has 130nm devices in production and medical has 250nm devices. Both industries are starting to look at adopting 90nm technology. The medical industry might never adopt 45nm due to reliability and power concerns. Because of the safety-related concerns, both industries need a way to demonstrate/quantify resilience, if moving to new reliability methodologies.

The automotive industry demands high-reliability, long-life products. There is a requirement of 0 PPM. Participants agreed that absolute 0 errors is not a sensible goal, being neither obtainable or quantifiable. To do rational engineering, we must help people understand that the goal is to manage the error rate realizing that there will always be some residual rate of errors. Discussion around this point made it sound like 0.1 PPM might be acceptable--i.e. 0 is just the approximation when expressed as PPM. The electronics in cars are expected to last the lifetime of the car, which can easily be 20+ years. The probability of dangerous failure per hour must be less than $10^{-7}$. They stated that they always need more performance.

The medical industry is primarily driven by low-power needs, as implantable devices must last 5-10 years on the same battery. Much like automobiles, compute processing needs continue to grow. They are starting to see an increase in soft errors in these devices. Soft errors are now in the PPM range. One particular failure manifestation for these devices is a power-on-reset (POR), where memory errors interact with device programming and lead to performance degradation conditions such as premature battery depletion. System response to these events is unpredictable, and in some cases can result in removal from the patient (explant). However, since soft-errors occur randomly and do not permanently damage the device, a new device is likely to see soft errors at the same rate as the old device.

Unlike other constituency groups, this group highlighted the need for better reliability than current silicon devices. Like consumer electronics and aerospace, this group is also looking at how analog devices and passives affect the entire system reliability.

# Infrastructure

The infrastructure group also briefed for the first time (slides). This group specifically discussed how the physical distribution of the sites affects reliability. As the power grid spans the entire country, access to system maintenance can be delayed by physical distance and there can be a delay in information propagation in the system. Once systems are deployed they are seldom removed from service, which means that the infrastructure systems is extremely heterogeneous and individual computing systems may span several generations of electronics. This heterogeneity necessitates flexible and adaptive reliability solutions that can be adopted to legacy systems that cannot be replaced.

Furthermore, the cost of computing does not dominate the system, as the computing systems are much cheaper than the machine they are controlling. With their highly distributed sites, they were particularly interested in autonomous remote monitoring of their systems; in cases where the elements of the system were not all directly powered (e.g. pipeline), there was interest in remote monitors that could scavenge their own power (e.g. harvest energy from the pressure of gas in a pipeline).

The infrastructure group discussed the value of degraded fallback. The aerospace group resonated with the value of degraded fallback and came up with their own slogan of "graceful degradation instead of abject failure." The infrastructure group also stated that their standard metric was availability instead of reliability.

# Roadmap

The roadmap group showed the results of their resilience study (slides) including the impact of increasing variation and aging effects on the rate of failure for key circuits (SRAM, latches, inverters). They have prepared draft for inclusion in the ITRS. They are continuing to work on adding the extrinsic noise model, which will shift the curves toward less reliable components.

# Metrics

The metrics group updated the meeting with their progress (slides). In this brief they discussed how the composition calculation must be more than summing FITs. If there are mitigation techniques, the summing would not take into account the benefit of the mitigation techniques. In that way a TMR-protected part would not have a FIT rate based on the parts (i.e., 3X), but a FIT rate reflecting the impact of this cross-checking system. They suggested decomposing the FIT rate calculation into persistent vs. transient errors and the impact of the error, such as detected and corrected (slowdown); failure of one application, virtual machine, or partition; full system failure; or silent data corruption (SDC). Infrastructure advocated that the availability metric should be two dimensional capturing both the event-rate and the time-to-recover for each type of error. The metrics group also discussed standardizing FIT metric ranges in a similar fashion as the infrastructure standards.

The metrics groups suggested that we measure the resilience of common electronics, such as house alarms, voting machined and point-of-sale terminals. They pointed out the reliability of these every day objects would be a concrete illustration of an academic topic that lay people might understand. They felt that this would be a step toward consumer reports or standards, making resilience a quantifiable selling point for products.

Their final suggestion was for a benchmark for reliability.

# Next Steps / Our path forward

- quotes from executives soon (mid. Nov.?)
- workshop summary (this) distributable by end of November
- DATE papers (driver for draft of key pieces) by end of November
- two-page executive summary during November (*original goal; looks like will be pushed out*)
- reports from constituency groups by Dec. 1
- input to SRC by December (this appears to be moving forward now)
- full report draft assembled in January? (cleaned up/polished in February?)
- lobbying SRC, DARPA, NSF, others? ...