

My Flings with Data Analysis

Venkatesh-Prasad Ranganath

Kansas State University

Dagstuhl Software Development Analytics Seminar

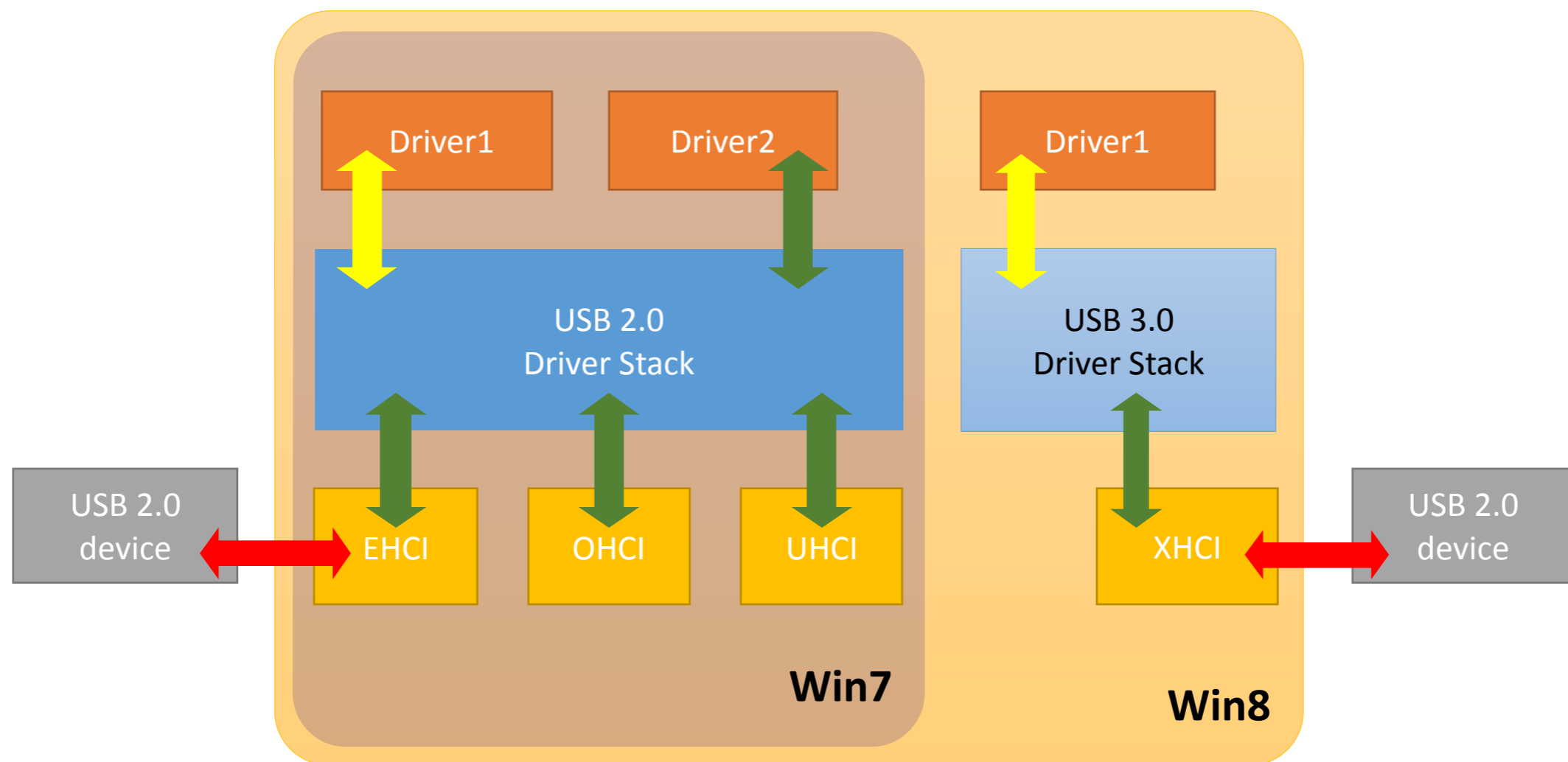
June 25, 2014

Observations

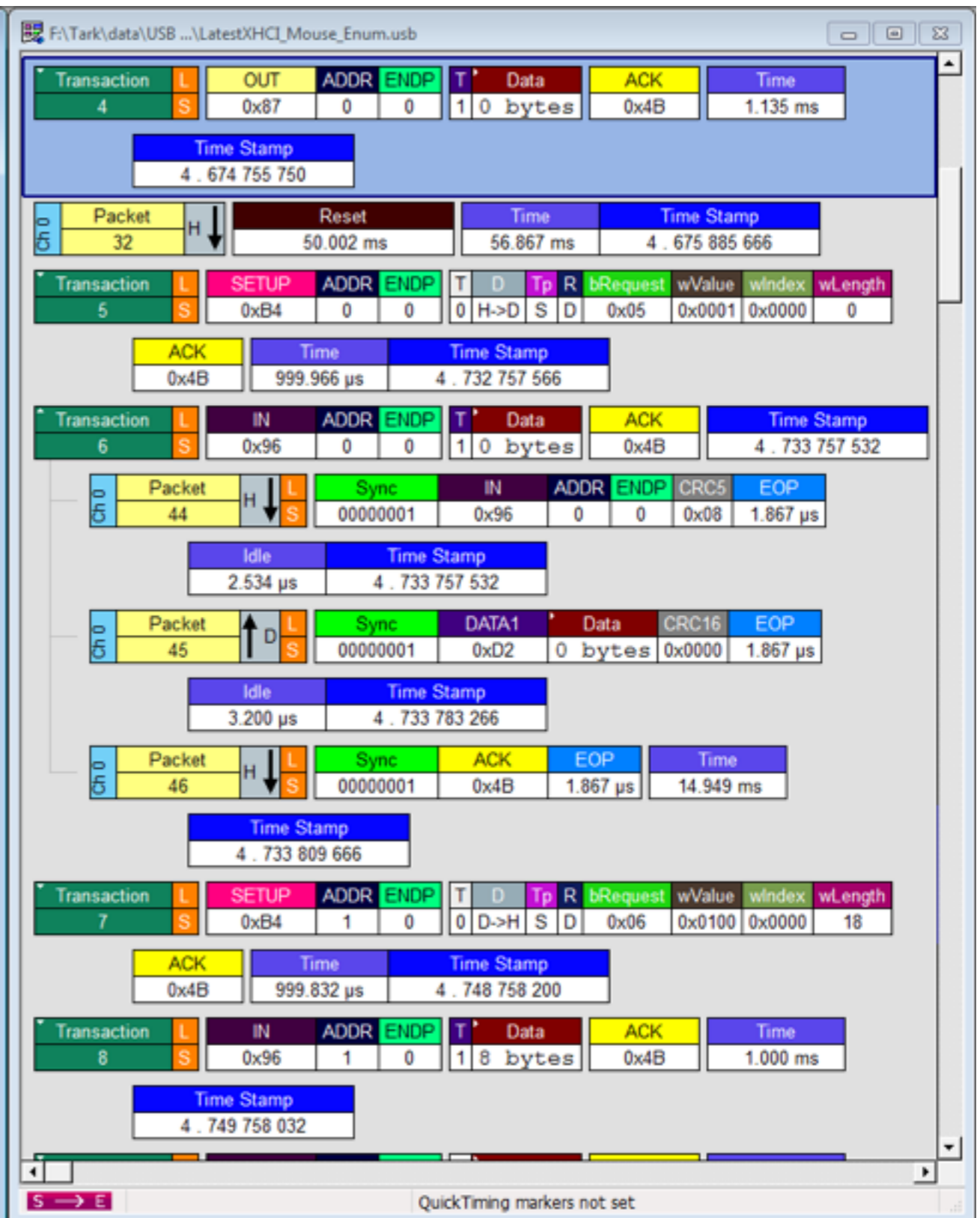
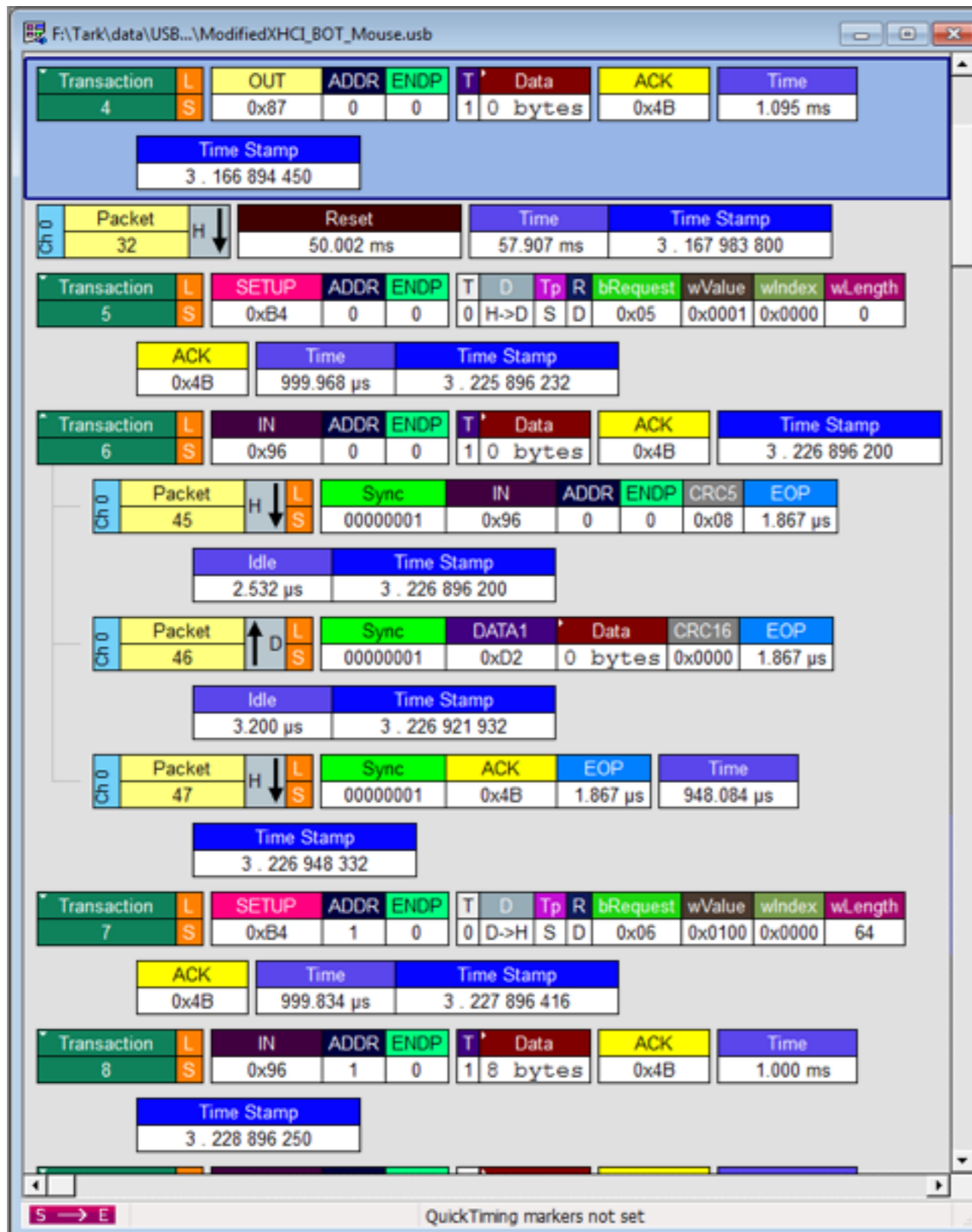
1. Involve the users early
2. Don't fear the unconventional
3. Embrace dynamic artifacts
 - Logs, core dumps, usage profiles, etc.
4. Have a feedback loop involving the users
5. Domain knowledge is a power pellet
6. Questions and answers first, methods later
7. Relevant features and simple methods work really well
8. Quick trumps slow and utility trumps quick
9. Automate (whenever possible)
10. Presentation matters
11. Models are good, simple explanations are better

Is USB 3.0/Win8 \approx USB 2.0/Win7?

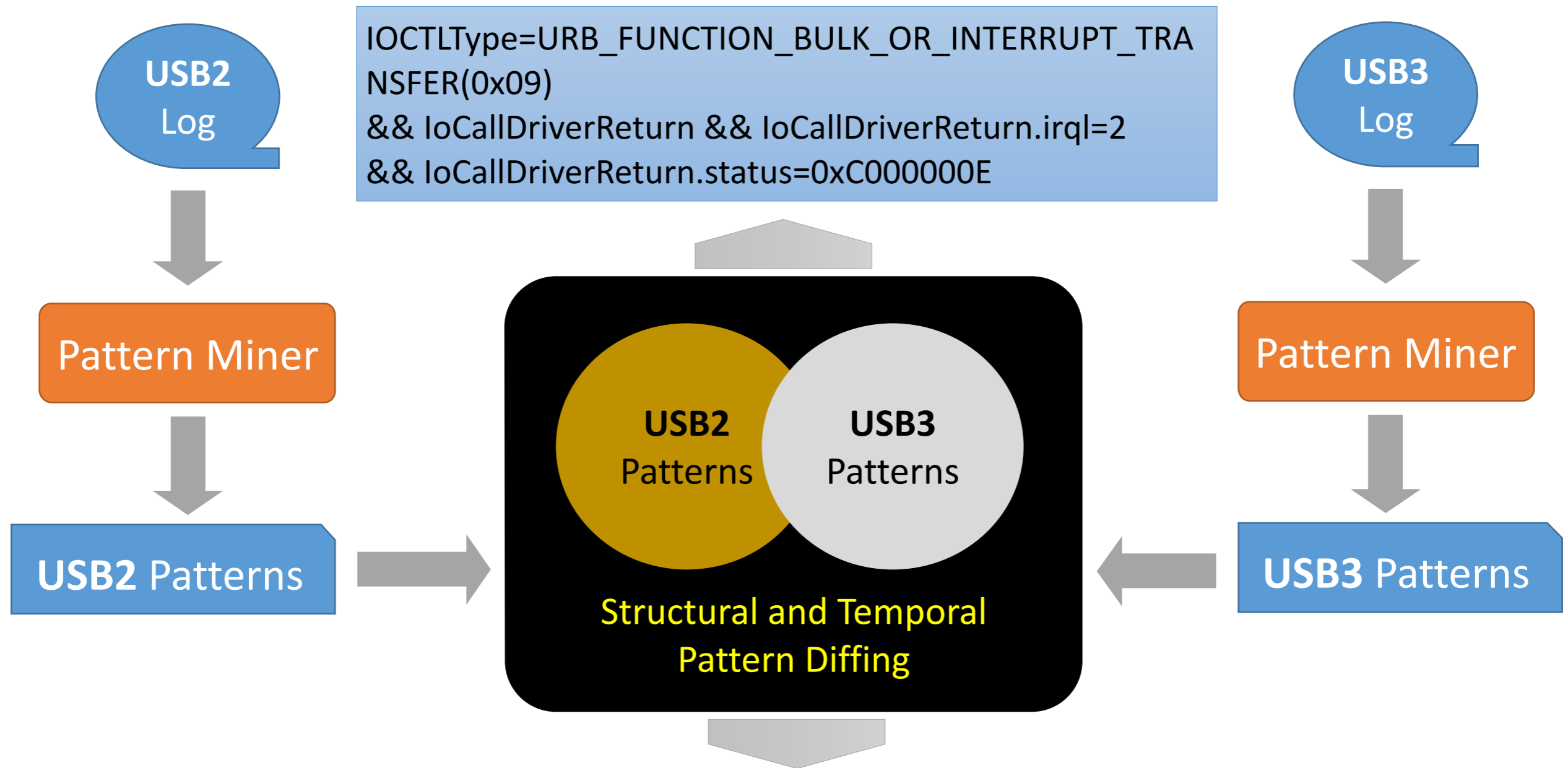
When a USB 2.0 device is plugged into a USB 3.0 port on Win8, the USB 3.0 stack in Win8 should behave as the USB 2.0 stack in Win7 (along both software and hardware interfaces).



One could



Patterns-based Compatibility Testing



DispatchIrp *forward alternates with* IrpCompletion && PreIoCompleteRequest when
IOCTLType=IRP_MJ_PNP(0x1B),IRP_MN_START_DEVICE(0x00), irpID=SAME, and
IrpSubmitDetails.irp.ioStackLocation.control=SAME

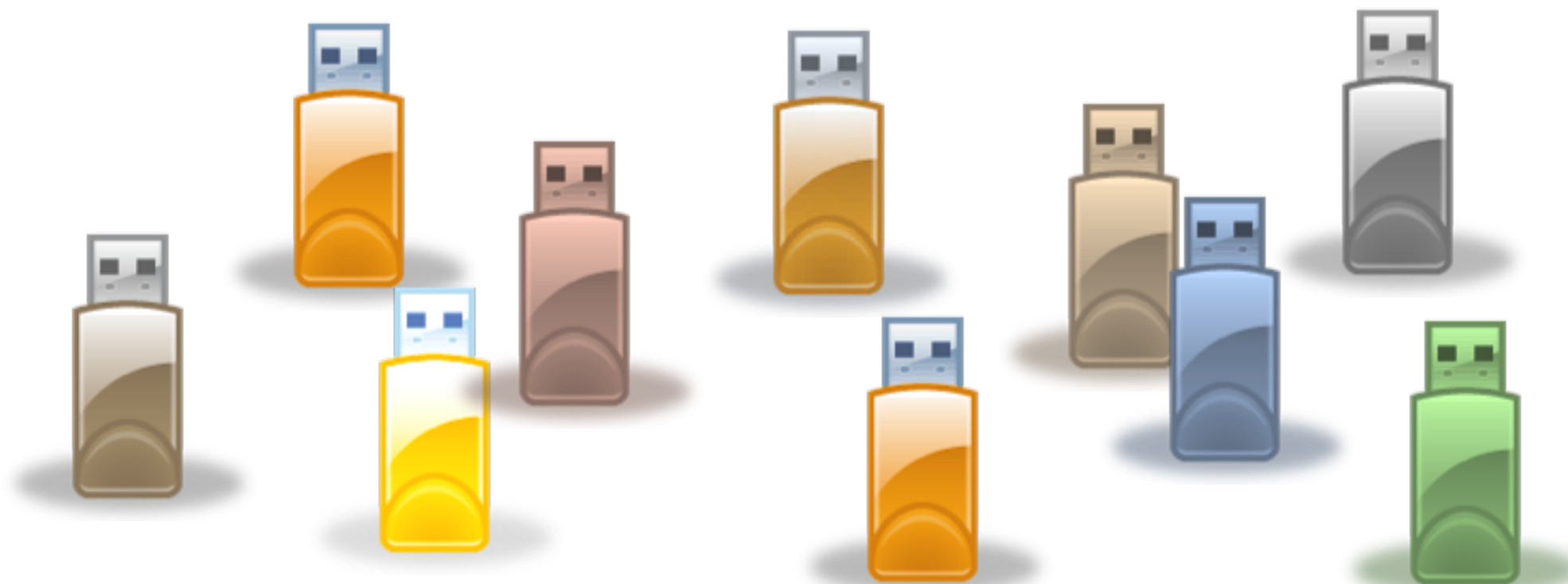
Results

Device Id	Detected	Reported	False +ve	Unique
1	9844	478	465	10
2	2545	15	11	2
3	743	4	1	1
4	1372	2	2	0
5	26118	55	55	0
6	26126	0	0	0
7	2320	0	0	0
8	27804	2	1	1
9	34985	115	98	4
10	51556	59	56	3
11	695	0	0	0
12	1372	0	0	0
13	3315	24	23	1
14	9299	3	0	2

Test Suite Reduction



Experience-based Selection



Patterns-based Test Suite Reduction



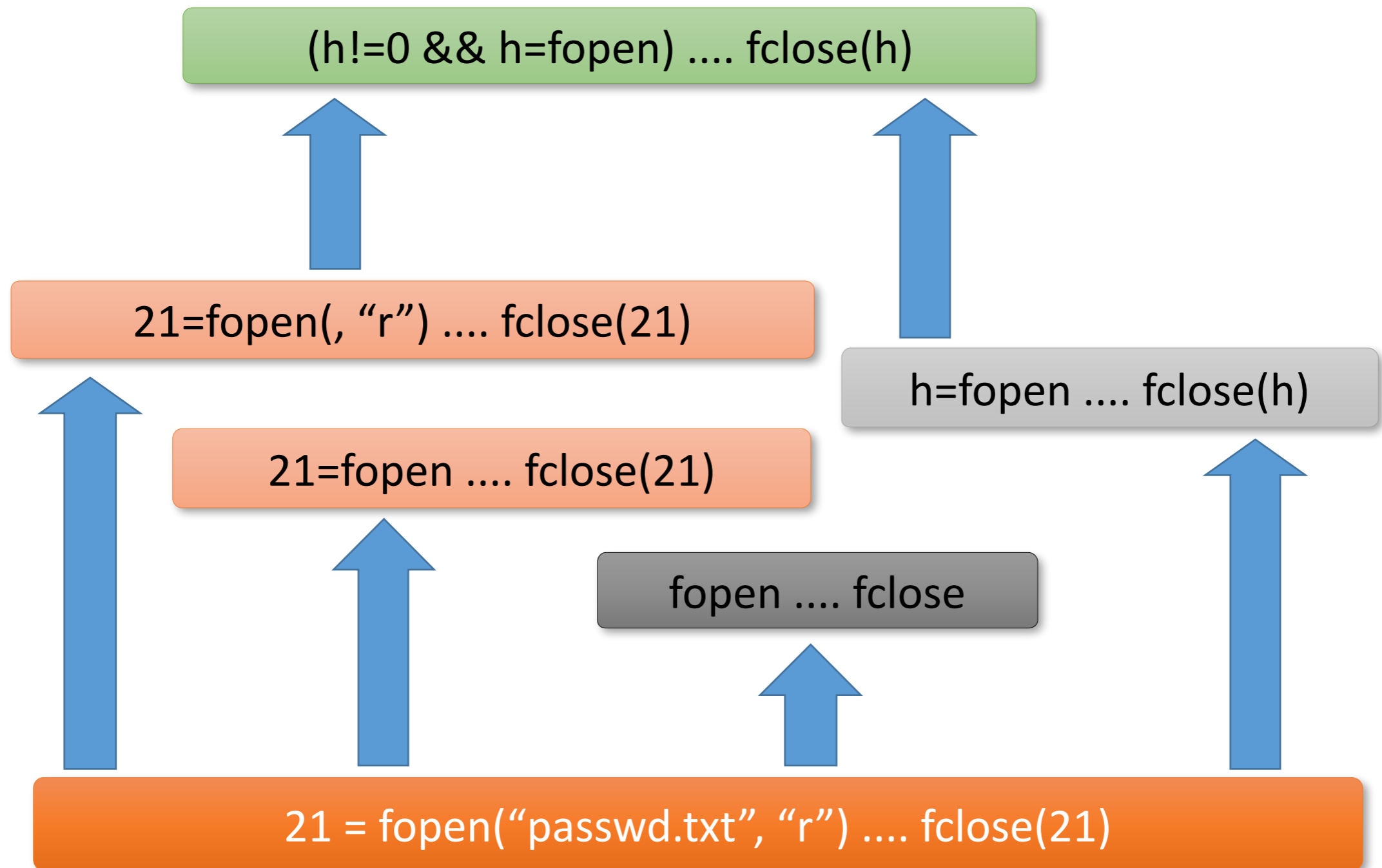
Cluster-n-Select



Results

- 50% reduction in test suite size
- 75-80% bugs uncovered (v/s human-based baseline)
- Fully automated (except for threshold setting)
- Flying without safety net

Structural & Temporal Patterns (w/ Data flow)



Observations

1. Involve the users early
2. Don't fear the unconventional
3. Embrace dynamic artifacts
 - Logs, core dumps, usage profiles, etc.
4. Have a feedback loop involving the users
5. Domain knowledge is a power pellet
6. Questions and answers first, methods later
7. Relevant features and simple methods work really well
8. Quick trumps slow and utility trumps quick
9. Automate (whenever possible)
10. Presentation matters
11. Models are good, simple explanations are better

Opportunities

- Combining of dynamic and static artifacts/techniques
 - Time to move beyond repositories
- Using data analysis to improve techniques
- Exploring/designing language for
 - Query
 - Visualization
- Scaling
 - Distributed Computing
 - GPUs

Questions

