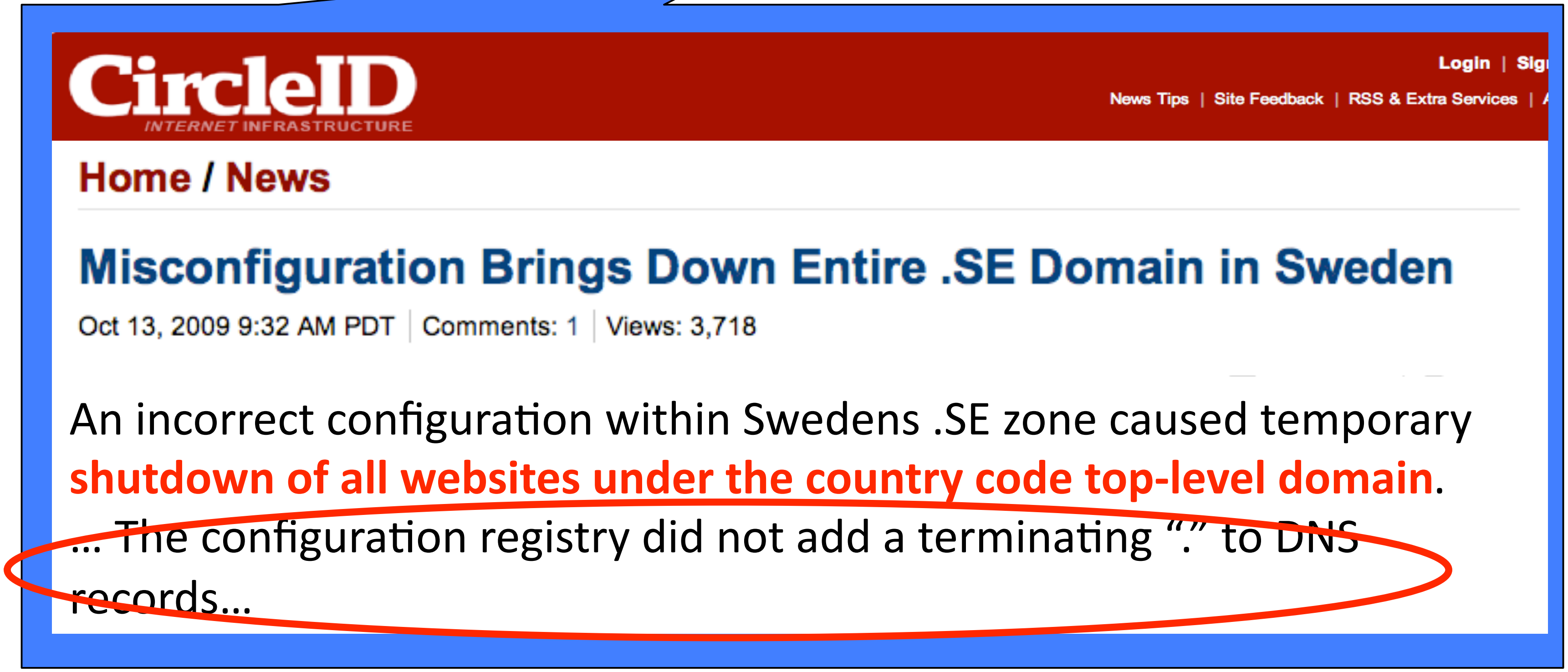
Context-based Online Configuration Error Detection

Ding Yuan[§], Yinglian Xie[¶], Rina Panigrahy[¶],
Junfeng Yang[⌈], Chad Verbowski[¶], Arunvijay Kumar[¶]

[¶]Microsoft Research, [§]UIUC and UCSD, [⌈]Columbia University,

Motivation

- Configuration errors are caused by erroneous settings in the software system



Home / News

Misconfiguration Brings Down Entire .SE Domain in Sweden

Oct 13, 2009 9:32 AM PDT | Comments: 1 | Views: 3,718

An incorrect configuration within Swedens .SE zone caused temporary **shutdown of all websites under the country code top-level domain**. … The configuration registry did not add a terminating "." to DNS records…

Motivation

- Configuration errors are caused by erroneous settings in the software system
- Huge impact
- Configuration error is a major root cause of today's system failures
  - 25% - 50% of system outages are caused by configuration error [Gray85,Jiang09,Kandula09]
  - This percentage is likely increasing

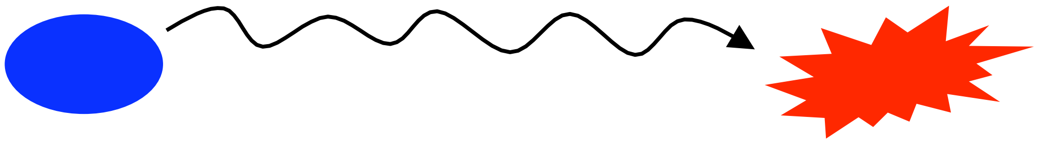
Existing Work

- Existing work focused on configuration error *diagnosis*
  - ConfAid[Attariyan10]
  - AutoBash[Su07]
  - Finding the Needle in the Haystack[Whitaker04]
  - PeerPressure [Wang04]
  - Self history constraint [Kiciman04]

Require manual error detection

Early Detection of Configuration Error

- Why we need early detection?



Configuration Error

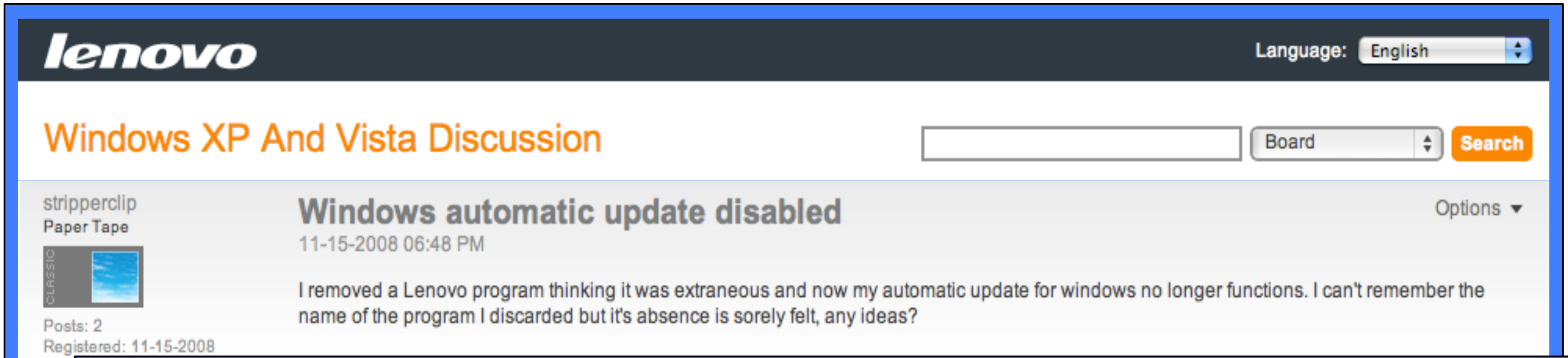
Failure

Windows Auto-Update disabled

Attacked by malware

  - Prevent error propagation
  - Hints for failure diagnosis
  - Especially useful in monitoring servers

Our goal: Automatically Detect Configuration Errors



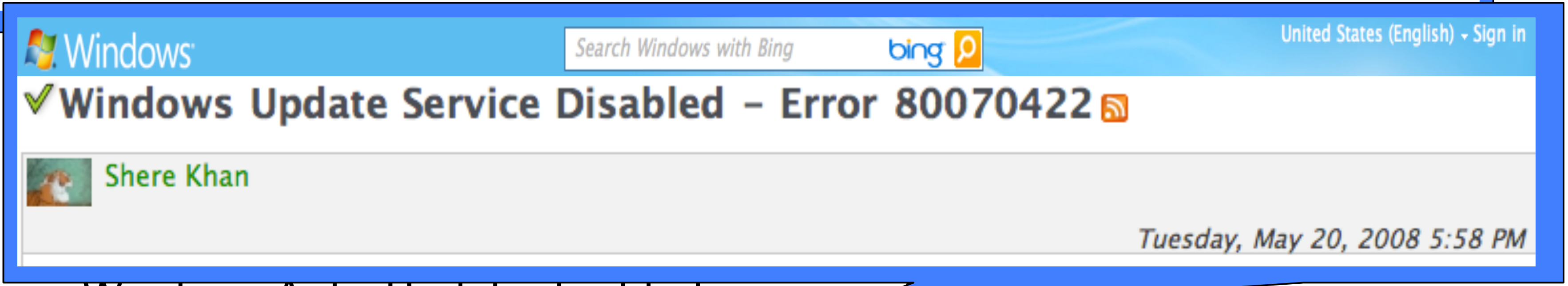
Windows XP And Vista Discussion

stripperclip
Paper Tape
Posts: 2
Registered: 11-15-2008

Windows automatic update disabled

11-15-2008 06:48 PM

I removed a Lenovo program thinking it was extraneous and now my automatic update for windows no longer functions. I can't remember the name of the program I discarded but it's absence is sorely felt, any ideas?



✓ Windows Update Service Disabled - Error 80070422

Shere Khan

Tuesday, May 20, 2008 5:58 PM

Windows Auto-Update disabled

attacked by malware

TechSupport FORUM

Security Alert

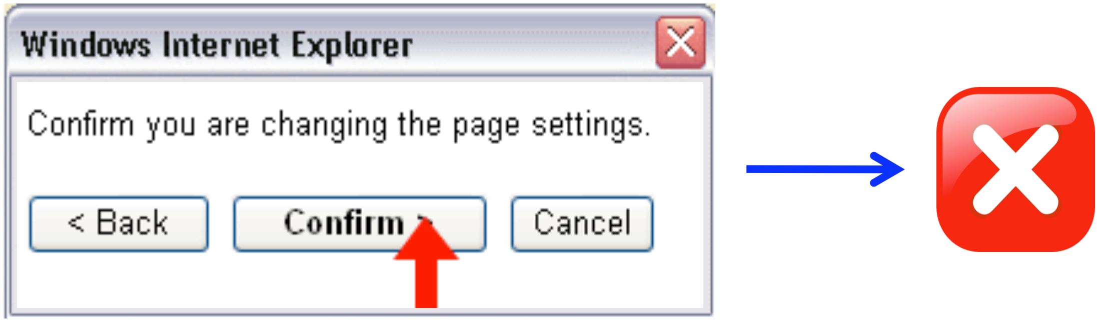
I am getting security alerts…

It looks like you might be having a malware

…Seems my Windows Update was disabled long ago…

Challenge

- First thought: report any configuration *change*
    - 10^4 writes/day per machine to Windows Registry
        - Majority are modifications to temporary Registry

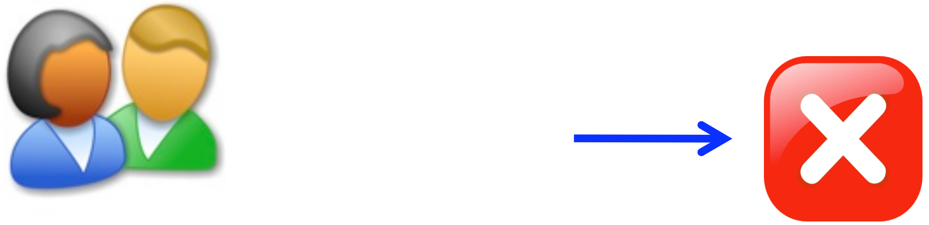


Windows Internet Explorer

Confirm you are changing the page settings.

Challenge

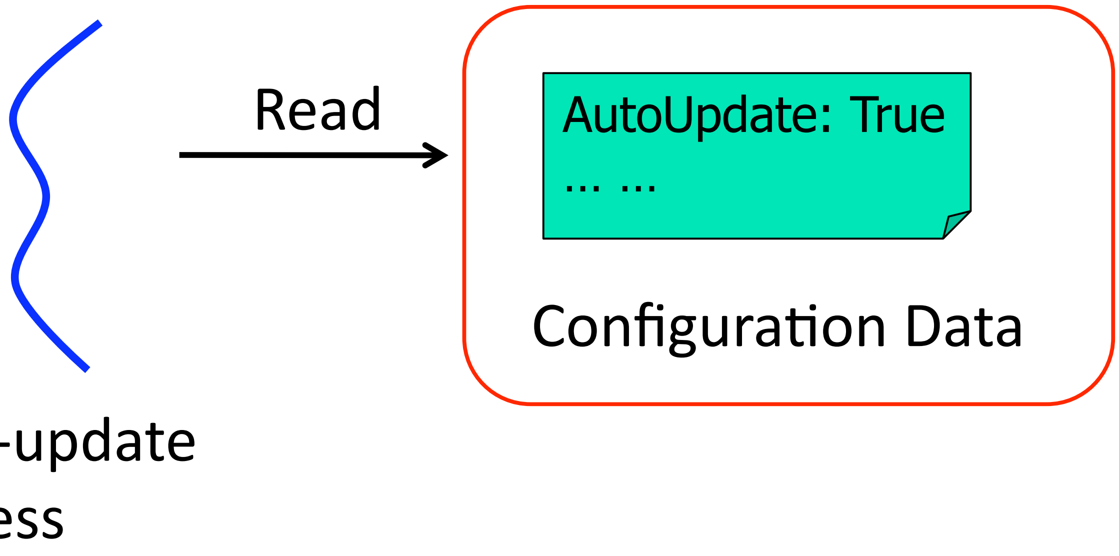
- First thought: report any configuration *change*
  - 10^4 writes/day per machine to Windows Registry
    - Majority are modifications to temporary Registry
- Only monitor the changes to ‘important’ configuration?
  - Too complicated: 200K Registry entries on single machine [WangOSDI04]



Change user previledge

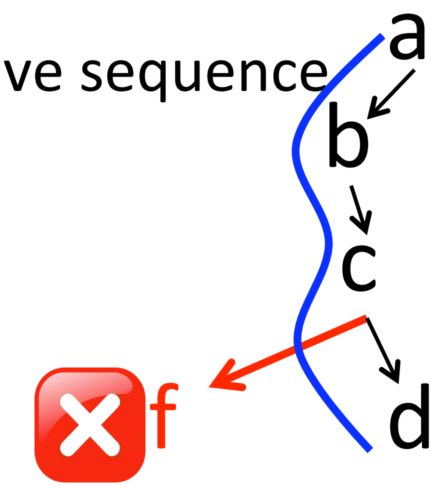
Our Observations

- Only those configurations that are *read* matter
  - Analyze read — configuration *access event*



Our Observations

- Only those configurations that are *read* matter
  - Analyze read — configuration *access event*
- Event sequences are repetitive and predictable
  - Externalize program's control flow
  - Report deviation from repetitive sequence



Contributions

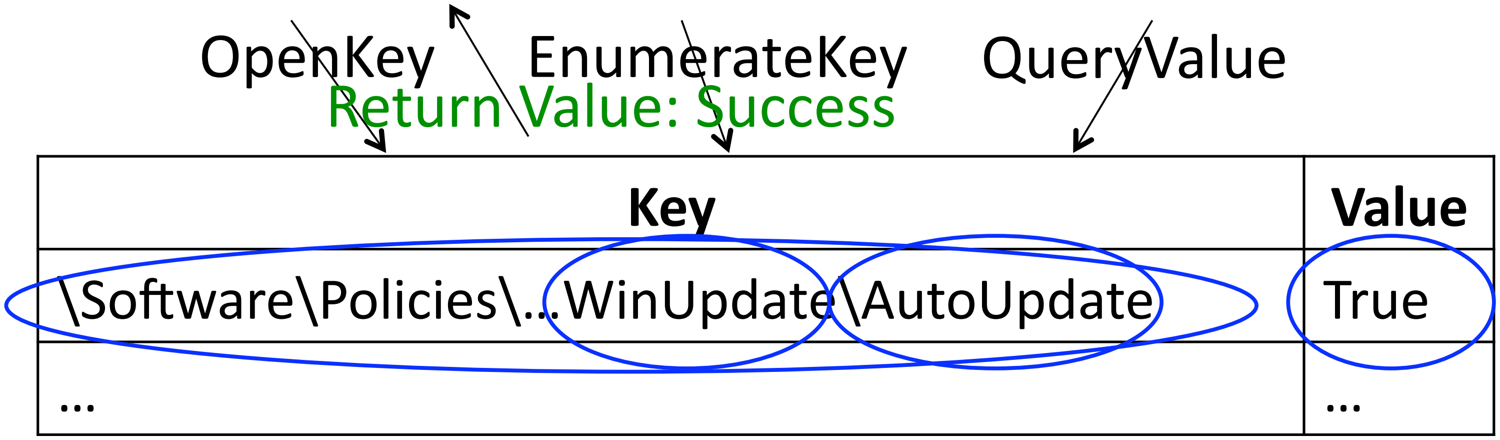
- CODE: online configuration error detection tool
  - Effective: detect configuration errors on-the-fly
  - Comprehensive: automatically monitor all the processes in OS (including kernel processes)
  - Reasonable false positive rate
  - Rich diagnostic information
  - Low overhead: < 1% CPU usage for 99% of time

Outline of the talk

- Motivations
- Background and Example
- Design and implementation
- Evaluation
- Related Work
- Limitations
- Conclusion

Windows Registry

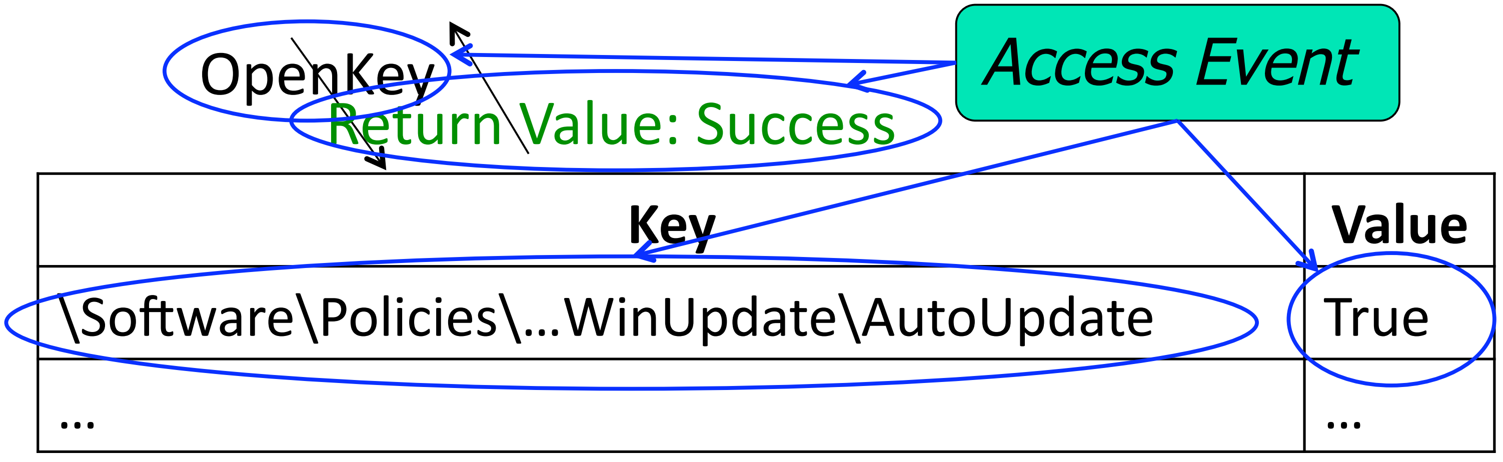
- Centralized configuration storage
  - Software, hardware and user settings
  - Key-Value pair
  - Standard interfaces for access Registry



Key	Value
\Software\Policies\...WinUpdate\AutoUpdate	True
...	...

Windows Registry

- Centralized configuration storage
  - Software, hardware and user settings
  - Key-Value pair
  - Standard interfaces for access Registry



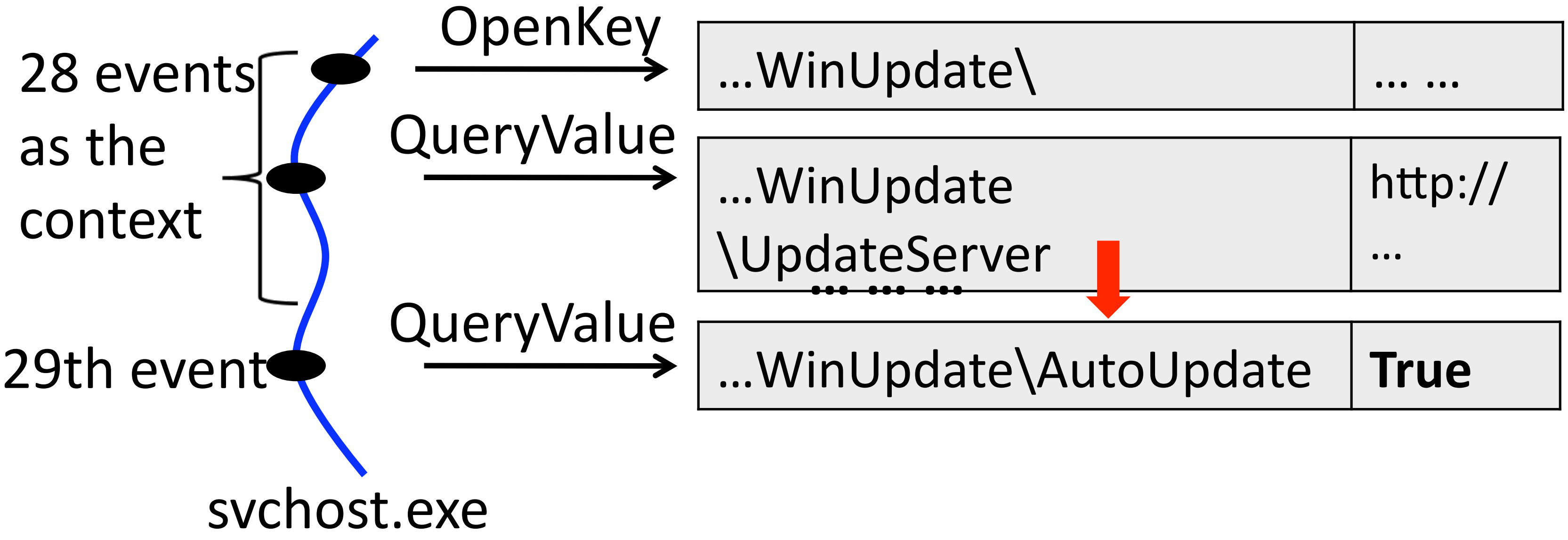
Access Event

OpenKey

Return Value: Success

Key	Value
\Software\Policies\…WinUpdate\AutoUpdate	True
…	…

Auto-Update Example



28 events as the context

29th event

svchost.exe

OpenKey

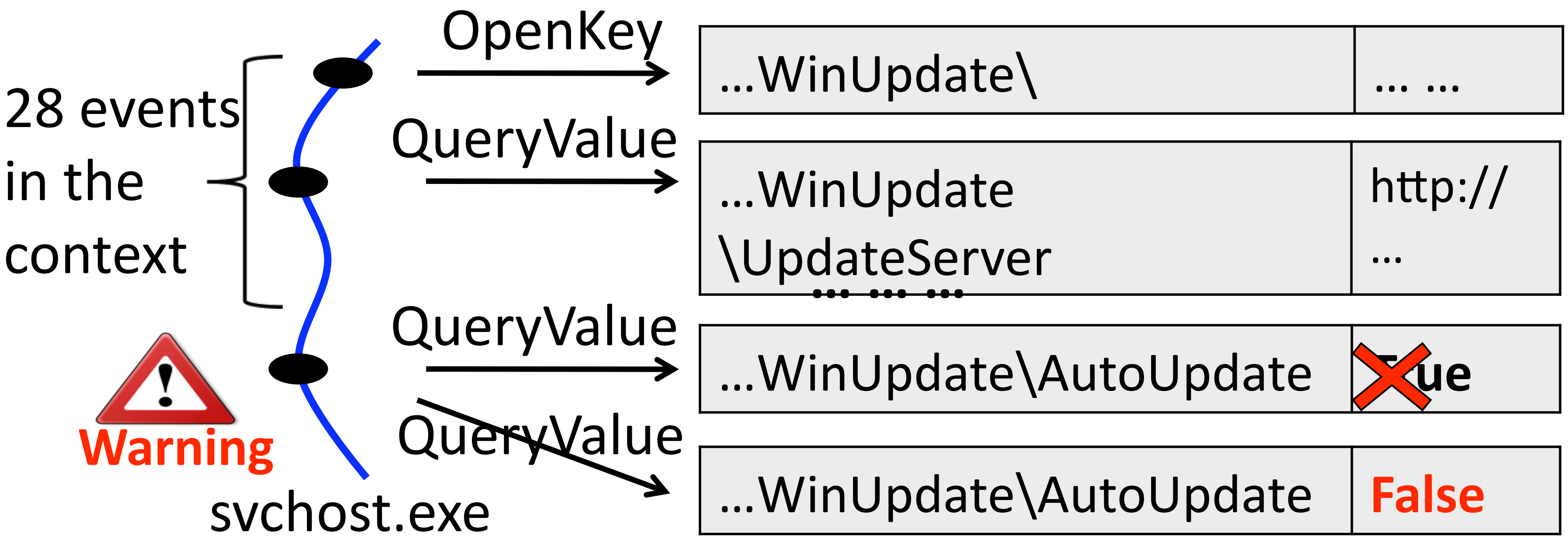
QueryValue

QueryValue

…WinUpdate\	… …
…WinUpdate \UpdateServer	http:// …
… … …	
…WinUpdate\AutoUpdate	**True**

Periodically checks for Windows update.

Auto-Update Example – Error case



…WinUpdate\	… …
…WinUpdate \UpdateServer	http:// …
…WinUpdate\AutoUpdate	~~True~~
…WinUpdate\AutoUpdate	**False**

Only when the modified Registry entry is read!

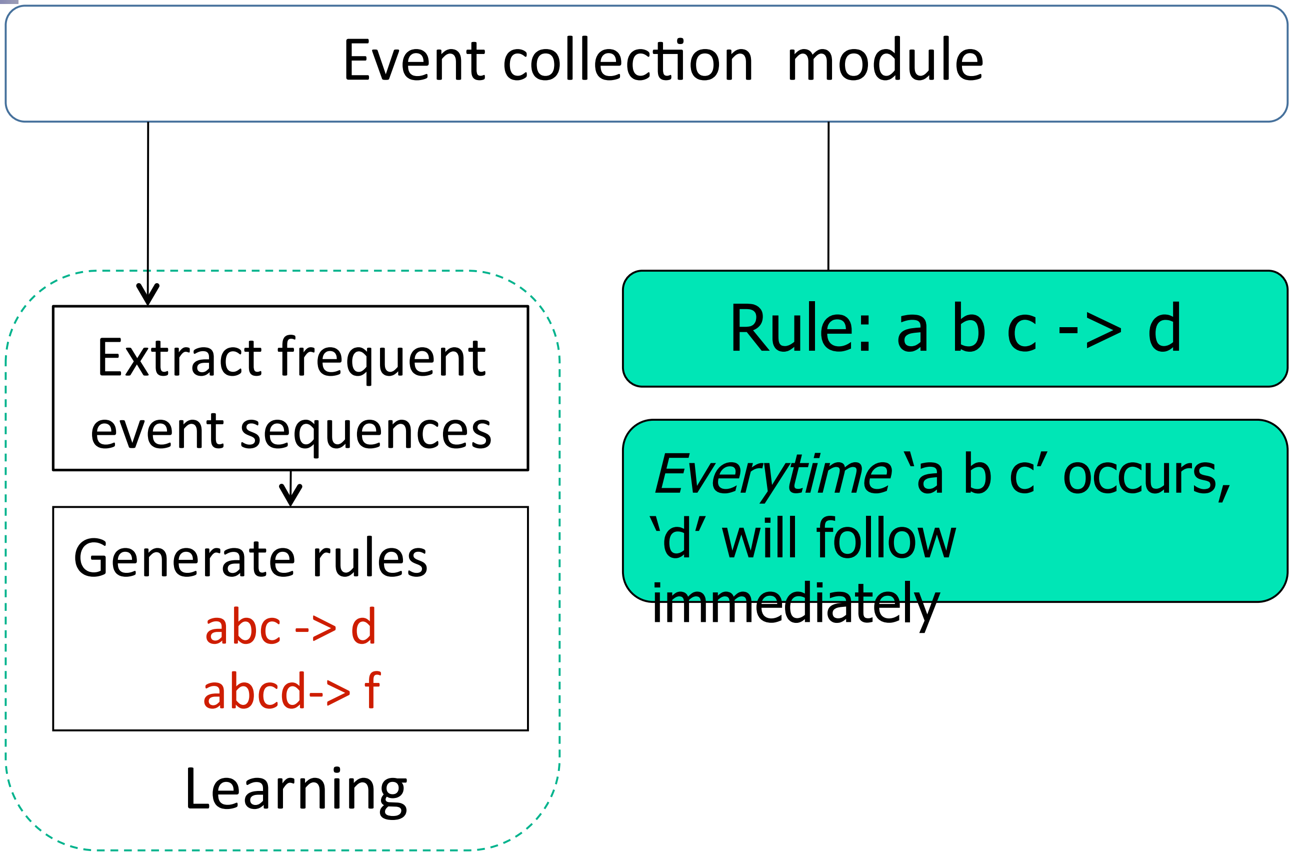
Expected: AutoUpdate = True

Observed: AutoUpdate = False

Modified by: explore.exe, at 2:03 PM, 4/6/2011

… …

Design Overview



Event collection module

Extract frequent event sequences

Generate rules

abc -> d

abcd-> f

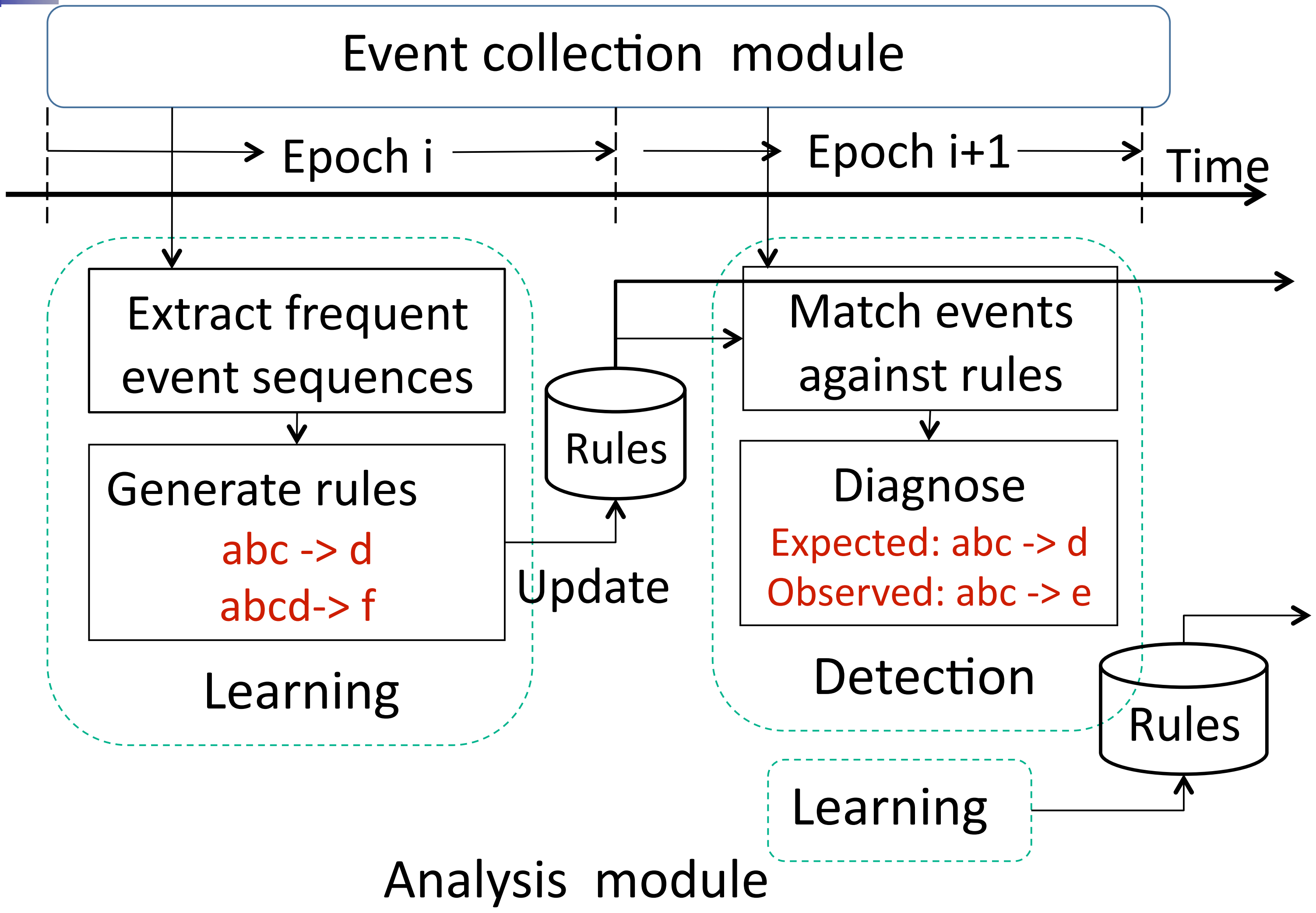
Learning

Rule: a b c -> d

Everytime 'a b c' occurs, 'd' will follow immediately

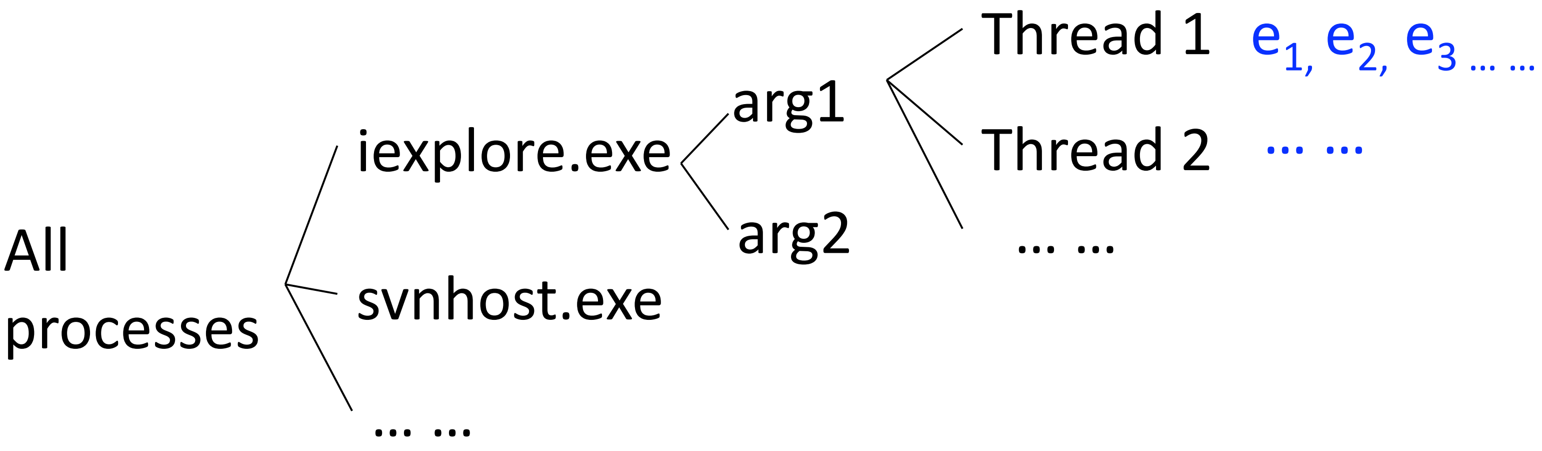
Analysis module

Design Overview



Event Collection

- Monitor the configuration access events
  - Sequences faithful to the program's control flow
  - Based on FDR [Verbowski08]
  - Negligible runtime & space overhead



All processes

iexplore.exe

svnhost.exe

… …

arg1

arg2

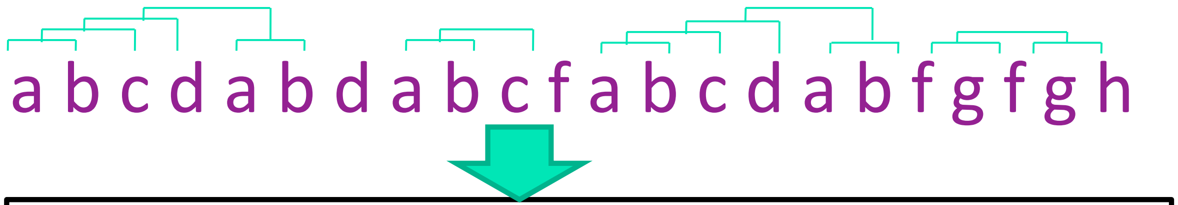
Thread 1 $e_{1,}$ $e_{2,}$ e_3 … …

Thread 2 … …

… …

Learn the frequent sequences

- Frequent Sequence Mining
  - **Efficiency:** streaming based method
- Sequitur algorithm [Manning97]
  - Streaming algorithm
  - Flexible pattern length



a b c d a b d a b c f a b c d a b f g f g h

R_1: a b -- 5 times

R_2: a b c d – 2 times

R_3: a b c d a b – 2 times

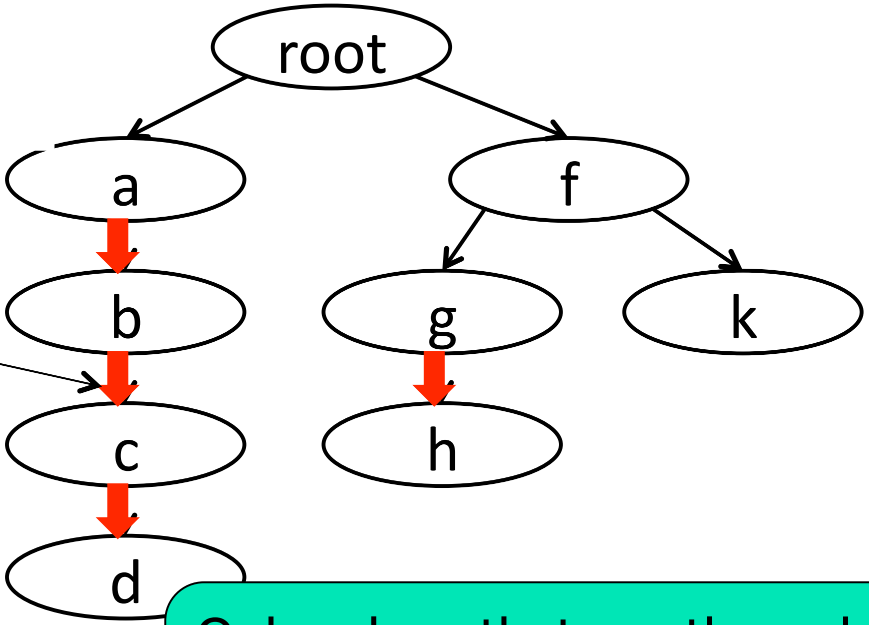
Deriving *Context -> Event* rules

- Put every frequent sequence into a prefix tree

Sequence 1: a b c d

Sequence 2: f g h

Sequence 3: f k



Represents 'ab -> c'

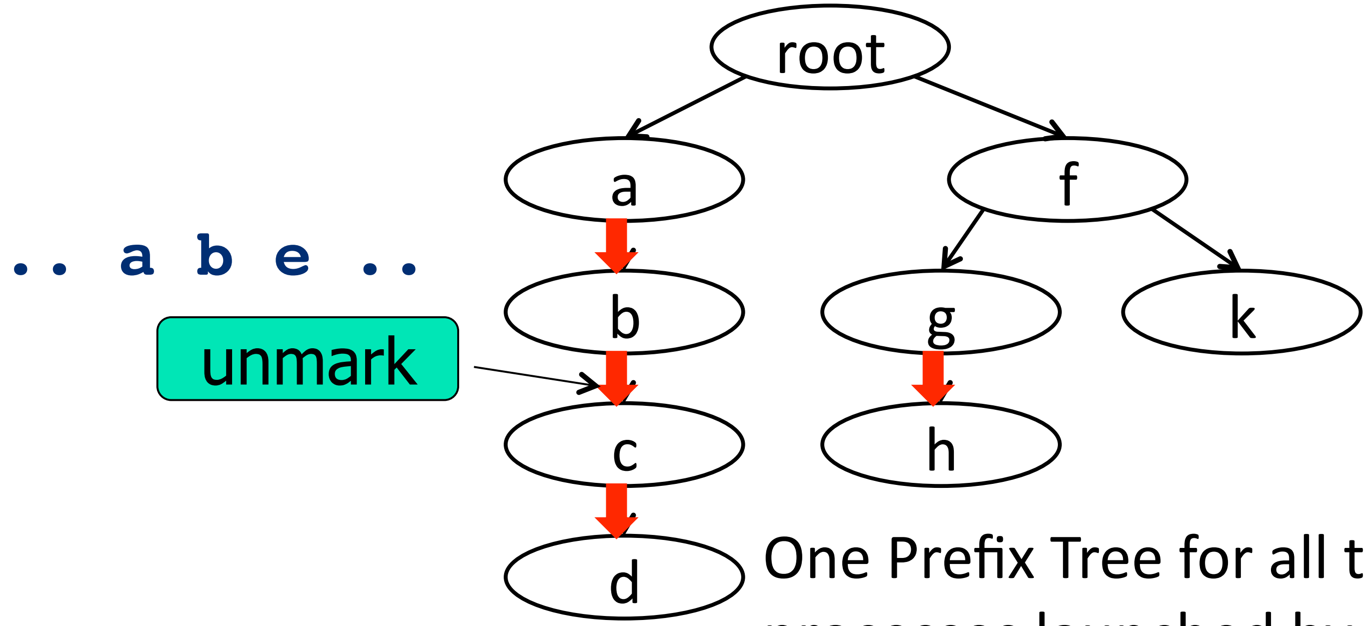
Each node is an event

Each edge might represent a rule

Only edges that are the only outgoing edge from the origin node are candidates to represent a rule

Deriving *Context -> Event* rules

- Not every candidate edge represents a rule

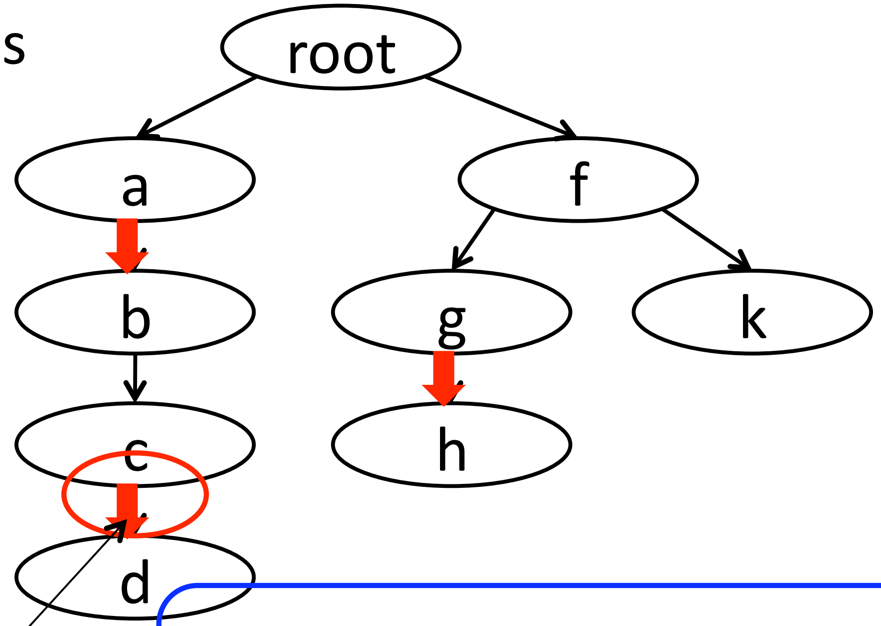


One Prefix Tree for all the processes launched by the same
process name and argument

Error Detection

- Report rule edge violation
  - Match incoming events against prefix tree

.. a b c e ..



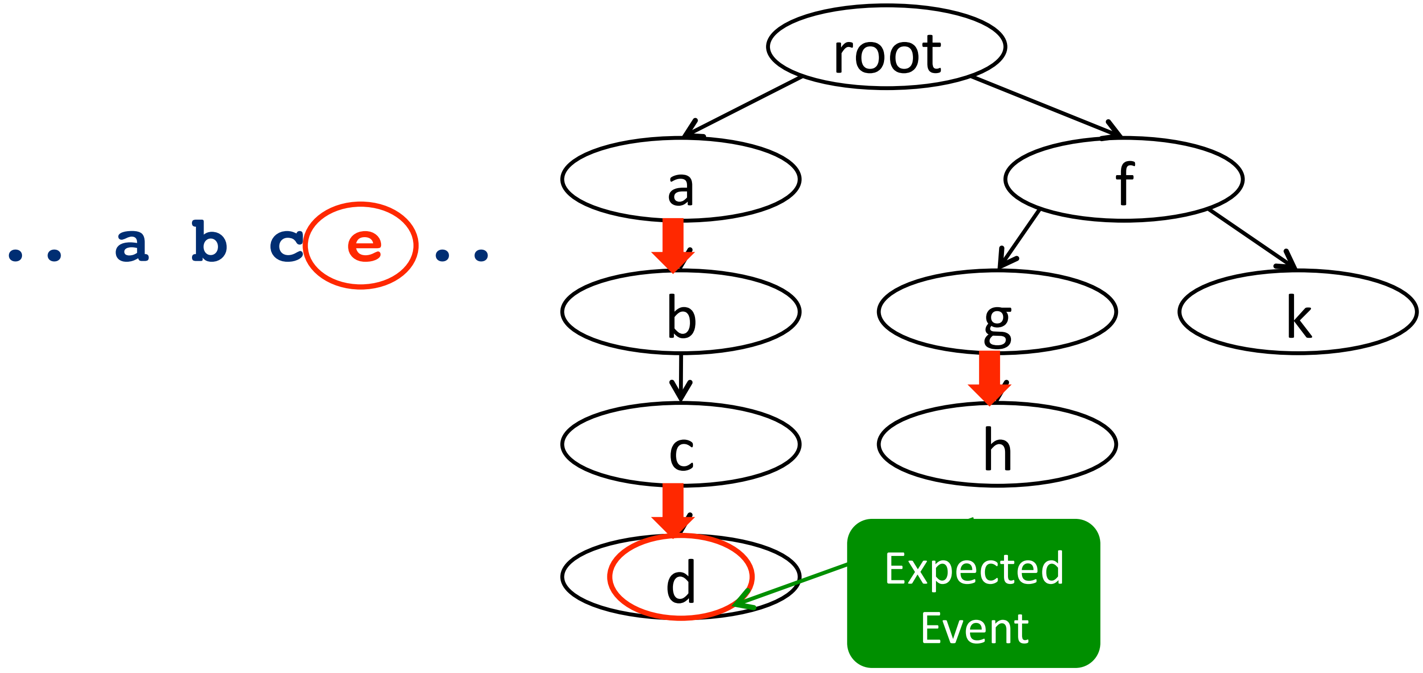
Report an error!

Represents ‘abc -> d’

A few heuristics to suppress false positives

Diagnostic Information

- What is the expected event
  - Help to recover from the error

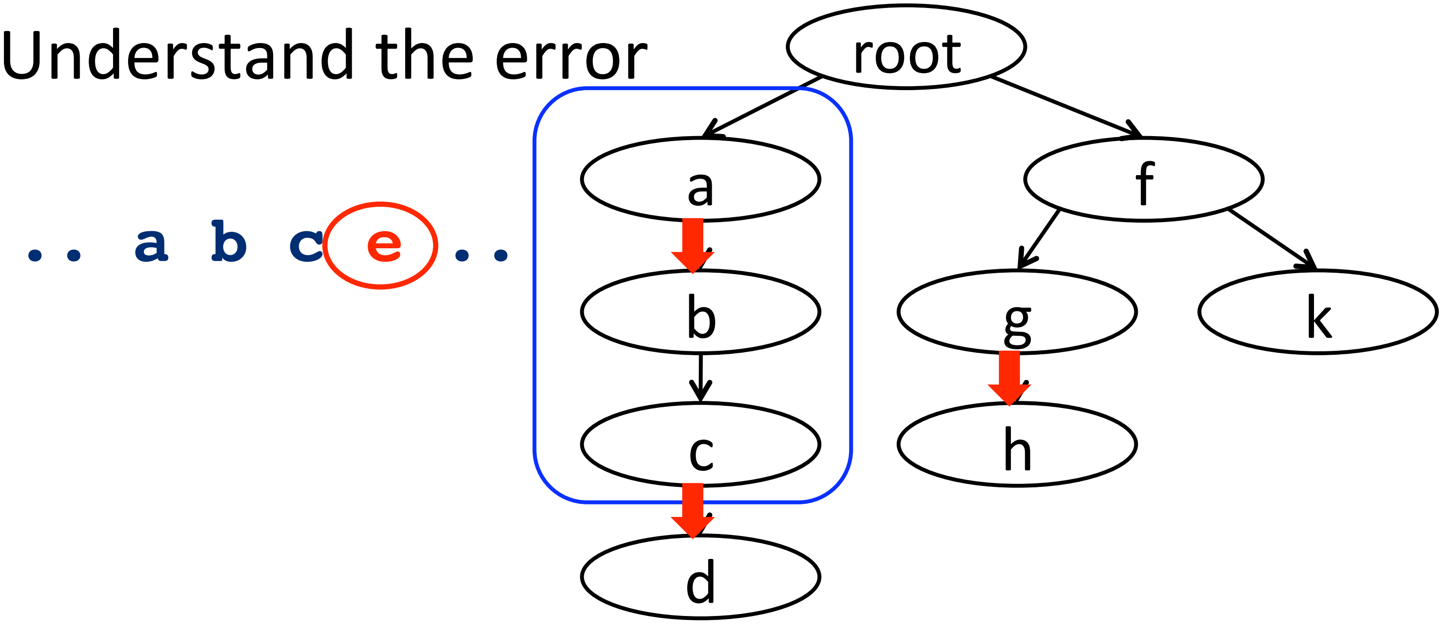


. . a b c e . .

Expected Event

Diagnostic Information

- What is the expected event
  - Help to recover from the error
- The context of the violation
  - Understand the error



Diagnostic Information

- What is the expected event
  - Help to recover from the error
- The context of the violation
- Which process modified the Registry that caused the error? And when?
  - Write buffer
- Examine the side effect of rolling back the Registry to its old data
  - All the other rules involving the new Registry data

Evaluation methodology

- False negative rate
  - Real configuration errors
  - Error injection
- False positive rate
  - Deployed on 10 actively using desktops and a server cluster with 8 servers running
- Performance

How many real world errors do we catch?

	Error Description	machines reproduced	# of cases detected
1	explorer-double-click	5	5
2	ie-advanceoptions	5	5
3	ie-search	2	2
4	ie-smbrandbitmap	1	1
5	ie-brandbitmap	1	1
6	ie-title	5	5
7	explorer-policy	5	5
8	explorer-shortcut	5	5
9	ie-password	4	4
10	ie-workoffline	5	4
11	outlook-emptytrash	4	4
Total:		42	41

Missing only 1 out of 42

Exhaustive Registry Corruption

- Exhaustively corrupted every Registry Key frequently accessed by Internet Explorer
    - Among 387 successfully corrupted Keys, CODE detected 374 (**97%**) of them
- CODE can effectively detect most of the Registry related configuration errors

False Positive Rate

- Deployed on 10 actively used desktop machines, 8 production servers
  - Over 30 days
  - Includes 78 software updates

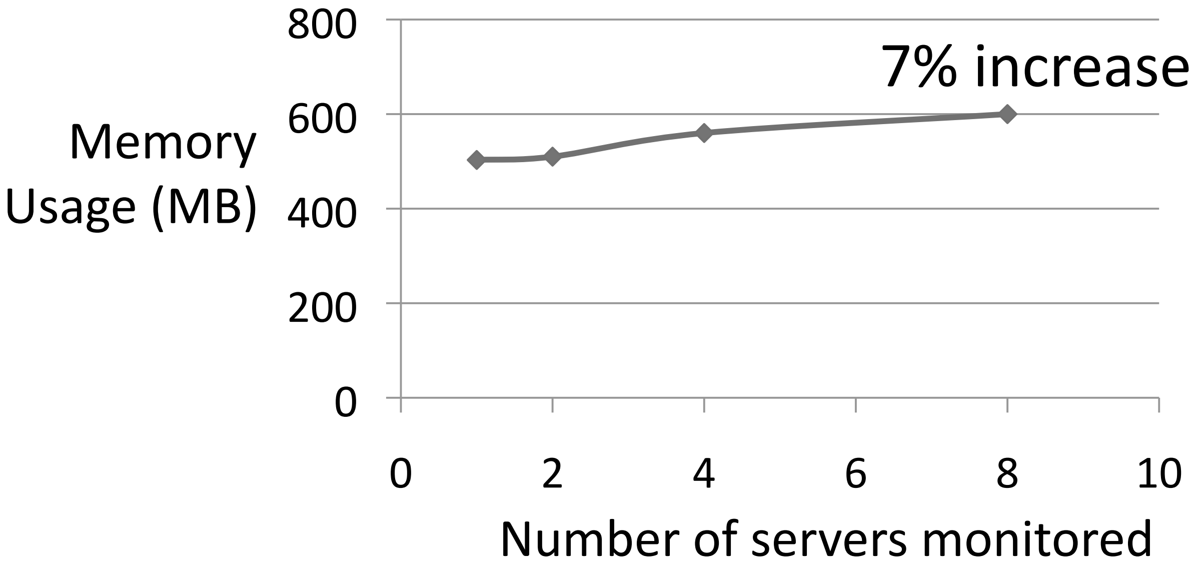
Warnings/ day	Average	Max	Min
Server	0.06	0.27	0
Desktop	0.26	0.96	0

Performance

- In all machines, CPU overhead is negligible
  - 1% over 99% of time
  - 10% - 25% peak usage

Performance

- In all machines, CPU overhead is negligible
- Memory Usage between 500MB – 900MB
- We can use one CODE process to monitor multiple servers with similar configuration setting



Related work

- Configuration error diagnosis
  - Key value pair based approaches [Wang04, Kiciman04]
  - Virtual Machine based [Whitaker04]
  - ConfAid[Attariyan10]
  - AutoBash[Su07]
- Sequence Analysis [Hofmeyr98,Wagner01]
  - Used in security
  - Different design
- Bug detection tools using symbolic execution
  - KLEE[OSDI08]

Limitations

- Cannot detect errors during installation
- Windows only
  - Key challenge on other systems: incercepting configuration accesses
- Still non-zero false positive rate
  - Limitation in truly differentiate user's rare intentional changes from errors

Conclusion

- CODE: Automatic online configuration error detection tool
  - Simple observation: key configuration access events form highly repetitive sequence
  - Effective and Efficient

Thanks

Top five causes for False Positives

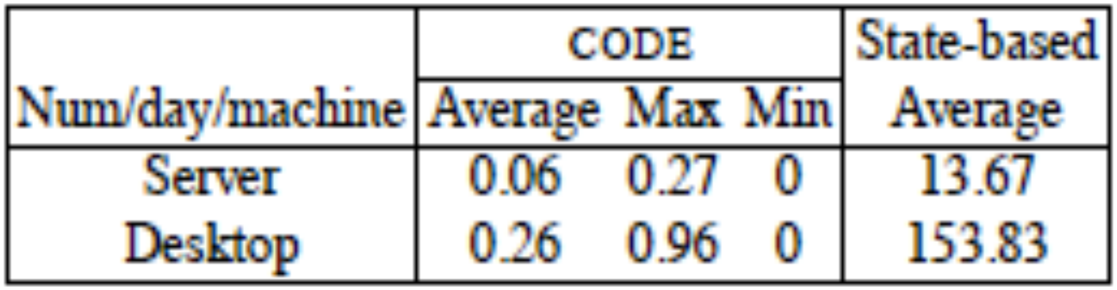
Name	Description	Percentage
File Association	The default program used to open different file types is changed.	24.1%
MRU List	Changes to most recently accessed files tracked by applications (e.g., explorer and IE)	12.7%
IE Cache	The meta-data for the IE Cache entities is changed.	3.8%
Session	The statistics for a user login session is updated	3.8%
Environment Variable	Environment Variable Changes	2.5%

Intentional configuration change that occurs infrequently

Impact of Software Updates

- During the month-long deployment on 10 desktops, only 5 warnings were due to software Updates (out of total 78)
  - 2 environment variable updates, one display icon update, one DLL update, one daylight saving time
- There was one most intrusive update
  - Office update from SP2 to SP3
  - 200 patches, modified 20,000 keys
  - Only 10 keys overlapped with CODE’s rule, causing only 1 warning

Comparison with state-based approach



	CODE			State-based
Num/day/machine	Average	Max	Min	Average
Server	0.06	0.27	0	13.67
Desktop	0.26	0.96	0	153.83