

# Automatic Misconfiguration Troubleshooting with *PeerPressure*

Helen J.Wang, John C.Platt, Yu Chen, Ruyun Zhang, Yi-Min Wang  
Microsoft Research

Presented by: Sriram Alwan

# Introduction

---

- Large amount of technical support time spent on troubleshooting.
- Many troubleshooting cases are due to misconfigurations.
- Misconfigurations caused by data in shared persistent stores like the Windows registry and UNIX resource files.
- Authors focus on Windows registry.
- Windows registry – provides hierarchical persistent storage for named, typed entries.

# Introduction

---

- Causes of misconfigurations?
  - Seemingly innocuous changes to shared system configurations
  - Software bugs
  - Security patches might introduce incompatible registry settings
  - Failed installation of applications
  - Manual intervention

# Introduction

---

- “Golden state” – refers to a perfect configuration
- Goals of a troubleshooting system
  - Effectiveness – System should identify a small set of sick configuration candidates within a short span of time
  - Automation – manual steps should be as minimum as possible

# Approach

---

- Compare the ‘sick’ machine with a ‘healthy’ machine
  - Difficult to identify a healthy machine automatically
- “Golden state is in the mass” – Application functions correctly on most of the machines; use statistics from a large enough set to obtain the *statistical golden state*. Combine with Bayesian estimates to identify misconfigurations on sick machines.

# PeerPressure Architecture

Captures *misconfiguration* suspects



App Tracer

Registry Entry Suspects

Entry	Data
HKLM\Software\Wubi...	On
HKLM\System\Setup...	0
HKCUR\Software\...	null

Turns user or machine-specific entries into *canonicalized* form

Canonicalizer

Search & Fetch

Statistical Analyzer

Peer-to-Peer Troubleshooting Community

GeneBank

Troubleshooting Result

Entry	Prob.
HKLM\Software\Wubi...	0.8
HKLM\System\Setup...	0.2
HKCUR\Software\...	0.008

PeerPressure

Database containing a

Uses Bayesian estimation to calculate probability of each suspect being sick. Output a *ranking report* based on the probabilities

number of machine snapshots

# PeerPressure Architecture

---

- Trial-and-Error fixing – User confirms whether sickness is cured
- Manual steps involved:
  - User runs sick application to record the suspects
  - User determines if sickness is cured
  - “Manual steps involve only the troubleshooting user and not any second parties”

# PeerPressure Algorithm: Intuition and Objectives

	Registry Key	Suspect Machine Registry Value	Machine 1 Value	Machine 2 Value	Machine 3 Value	Machine 4 Value	Machine 5 Value
e1	.jpg/contentType	<i>image/jpeg</i>	image/jpeg	image/jpeg	image/jpeg	image/jpeg	image/jpeg
e2	.htc/contentType	not exist	text/x-comp	text/x-comp	text/x-comp	text/x-comp	text/x-comp
e3	url-visited	yahoo	hotmail	nytimes	SFGate	google	friendster

e1: Probably Healthy

e2: most probably sick

e3: can't say due to "natural biological diversity"

Types of state in canonicalized entries:

1. Application configuration states.
2. Operational states. Ex: timestamps, caches, window positions

# PeerPressure Algorithm: Formulation

$$P(S|V) = \frac{P(V|S)P(S)}{P(V|S)P(S) + P(V|H)P(H)} \quad (1)$$

$$P(S) = \frac{1}{t}, \quad P(H) = 1 - \frac{1}{t},$$

$$P(V|S) = \frac{1}{c}$$

$$P(V|H) = \frac{m}{N}, \quad (2)$$

$$P(S|V) = \frac{N}{N + cm(t-1)} \quad (3)$$

When  $m=0$ ;  $P(S|V)=1$  !!!

Bayesian estimation used to overcome this.

Vector  $p_j$ : The probability of an event happening and its outcome being the value  $V_j$

$p_j$  follows Dirichlet probability distribution.

$n_j$ : count vector – number of possible counts for a  $V_j$

$m_j$ : count of number of values matching suspect value

$$P(V_j|H) = \frac{m_j + n}{N + cn} \quad (4)$$

$$P(S|V) = \frac{N + cn}{N + cnt + cm(t-1)} \quad (5)$$

$N$	Number of sample machines
$t$	Number of suspect registry keys
$i$	The index for the suspect key (from 1 to $t$ )
$V_i$	The value of a suspect key $i$
$c$	Cardinality: the number of possible sample values for a suspect key
$m$	The number of samples that match the suspect value
$P(S)$	The prior probability that a suspect key is sick
$P(H)$	$1 - P(S)$
$P(S V)$	The probability that a suspect key is sick given its value
$P(V S)$	The probability that a sick suspect key $i$ has value $V_i$

# PeerPressure Algorithm: Asymptotic Analysis

---

$$\lim_{m \rightarrow \infty} P(S|V) = 0.$$

$$\lim_{n \rightarrow \infty} P(S|V) = \frac{1}{t} = P(S).$$

$$\lim_{N \rightarrow \infty} P(S|V) = \frac{1}{1 + cf(t-1)}.$$

$$\lim_{N \rightarrow 0} P(S|V) = \frac{1}{t} = P(S).$$

$$\lim_{c \rightarrow \infty, N \rightarrow \infty} P(S|V) = \lim_{c \rightarrow \infty} \frac{1}{1 + cf(t-1)} = 0.$$

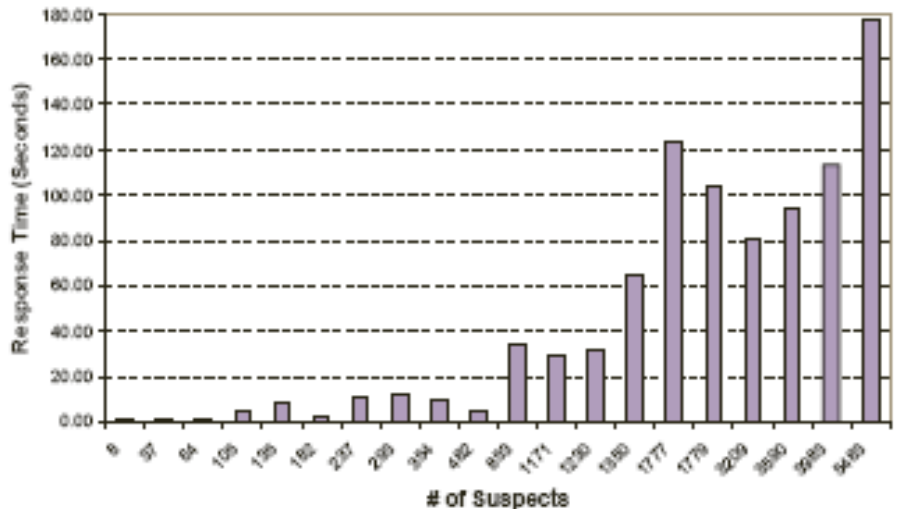
$$P(S|V) = \frac{N + cn}{N + tcn}.$$

$N$	Number of sample machines
$t$	Number of suspect registry keys
$i$	The index for the suspect key (from 1 to $t$ )
$V_i$	The value of a suspect key $i$
$c$	Cardinality: the number of possible sample values for a suspect key
$m$	The number of samples that match the suspect value
$P(S)$	The prior probability that a suspect key is sick
$P(H)$	$1 - P(S)$
$P(S V)$	The probability that a suspect key is sick given its value
$P(V S)$	The probability that a sick suspect key $i$ has value $V_i$

# PeerPressure Prototype

---

- Genebank database – Microsoft SQL Server 2000, contains snapshots from 87 Windows XP PCs.
- PeerPressure troubleshooter implemented in C#
- “Data Sanitization” – Unification of different representations of the same value
- Dual Intel Xeon 2.4 GHz CPU workstation with 1Gb RAM hosts SQL Server.



# Troubleshooting effectiveness: Registry Characteristics

---

Maximum registry size	333,193
Minimum registry size	77,517
Average registry size	198,376
Median registry size	198,608
Distinct canonicalized entries in GeneBank	1,476,665
Common canonicalized entries	43,913
Distinct entries data-sanitized	918,898

*Unseen* – Values that are unknown to the GeneBank

*No Entry* – Values that don't exist on certain machines

87% of registry entries – cardinality 2

94 % - no more than 3

97% - no more than 4

“Low cardinality contributes to excellent troubleshooting results of PeerPressure”

# PeerPressure Performance

- Real World Troubleshooting Cases used

ID	Name	Description of Problem
1	System Restore	No available checkpoints are displayed because the calendar control object cannot be started due to a missing Registry entry.
2	JPG	Right clicking on a JPG image and choosing the Send To → Mail Recipient option does not offer the resize option dialog box due to a missing Registry entry.
3	Outlook	User is always asked upon exiting Outlook whether she wants to permanently delete all emails in the Deleted Items folder, due to a hard-to-find setting.
4	IE Passwords	Internet Explorer (IE) browser does not offer to automatically save passwords; the option to re-enable the feature is difficult to find.
5	Media Player	Windows Media Player "Open Url" function fails if the EnableAutodial Registry entry is changed from 0 to 1 on a corporate desktop.
6	IM	MSN Instant Messenger (IM) significantly slows down if the firewall client is disabled on a corporate desktop.
7	IE Proxy	IE on a machine with a corporate proxy setting fails when the machine is connected to a home network.
8	IE Offline	IE "Work Offline" option may be automatically turned on without user knowledge; user is then presented with a cached offline page instead of the default start page when launching IE.
9	Taskbar	IE windows are unexpectedly grouped under the Windows Explorer taskbar group, due to the addition of a Registry entry.
10	Network Connections	Control Panel → Network Connections shows nothing, due to a missing Registry key.
11	Folder Double-Clicking	Double clicking any folder in the right pane of Windows Explorer incorrectly brings up the "Search Results" window.
12	Outlook Express	Microsoft Outlook could not be started because the Outlook Express installation appears to be missing, due to a missing Registry key.
13	Cannot Start Executables	Double-clicking any EXE file does not launch the application.
14	Shortcut	Double-clicking any shortcut does not launch the application.
15	IE Menu Bar	IE menu bar disappears due to a corrupted Registry key name.
16	IE Favorites	IE uses the "unknown file type icon" for some of the links in the Favorites.
17	Sound Problem	Warning sound is missing when an invalid command was typed into Start→Run.
18	IE New Window	Right-clicking a link inside IE and choosing "Open in New Window" shows nothing.
19	Yahoo Toolbar	Yahoo Companion per-user installation affects all users.
20	Media Player in IE	Internet Explorer always launches Media Player on the left pane.

# PeerPressure Performance

---

- Root Cause Ranking

Case	Rank	Ties	# of Suspects	Cardinality	# of Matches	# of Samples
1. System Restore	1	0	1350	3	1	87
2. JPG	16	0	1779	3	5	87
3. Outlook	1	0	37	4	7	566
4. IE Passwords	1	0	135	4	1	566
5. Media Player	1	0	182	6	1	566
6. IM	12	0	1777	4	8	87
7. IE Proxy	1	0	1171	16	0	566
8. IE Offline	1	0	1230	4	1	566
9. Taskbar	1	0	64	4	2	566
10. Network Connections	2	0	354	2	1	87
11. Folder Double-Click	2	1	26308	2	0	87
12. Outlook Express	3	0	482	2	0	87
13. Cannot Start Executables	1	0	237	2	0	87
14. Shortcut	1	0	105	2	0	87
15. IE Menu bar	1	2	3590	2	0	87
16. IE Favorites	2	0	3209	3	0	87
17. Sound Problem	1	0	8	1	0	566
18. IE New Window	1	0	853	2	0	87
19. Yahoo Tool bar	n/a					
20. MediaPlayer in IE	9	0	5483	65	0	566

# PeerPressure Performance

---

- False Positives
  - Large cardinality of root-cause entry
  - Relation between root-cause entry and other entries in the suspect set
  - GeneBank is not pristine

# PeerPressure Performance

I1

Case	5 Samples (Ties)	10 (Ties)	20 (Ties)	30 (Ties)	50 (Ties)	87 (Ties)	# of matches
<b>Perfect ranking regardless of the sample set size</b>							
5. Media Player	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1
14. Invalid Shortcut	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	0
17. Sound Problem	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	0
<b>Ranking Trend not solely dependent on the sample set size</b>							
10. Network Connections	1.6 (1)	1.4 (0.6)	2 (0.2)	1.4 (0.2)	1.4 (0)	2 (0)	1
20. Media Player in IE	6.2 (0.2)	6.2 (0)	8 (0)	11 (0)	11.2 (0)	9 (0)	0
2. JPG	8.4 (0.2)	13.4 (0.4)	14.6 (0.2)	13 (0.2)	14.2 (0)	16 (0)	5
6. IM	15.6 (1.6)	104 (0.2)	20 (0)	15.4 (0)	14.6 (0)	8 (0)	8
<b>Larger Sample Set improves ranking</b>							
8. IE Offline	1 (0.2)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	1
13. Cannot Start Executables	1 (0.4)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	0
1. System Restore	1 (0)	1 (0.2)	1 (0.2)	1 (0.2)	1 (0)	1 (0)	1
9. Taskbar	1.6 (5)	1 (0)	1 (0)	1 (0)	1 (0)	1 (0)	2
3. Outlook	2.2 (0.4)	1.4 (0.8)	1.6 (0.8)	1.4 (0.8)	1 (0)	1 (0)	7
4. IE Passwords	5.8 (8.2)	3.2 (2.4)	3.2 (2.4)	1 (0)	1 (0)	1 (0)	1
7. IE Proxy	3.4 (1.8)	2.2 (0.2)	2 (0.8)	3(3.2)	1(0)	1 (0)	0
15. IE No Menu Bar	6.4 (10.8)	3.2 (3.6)	2.2 (2.6)	1.6 (2.4)	1 (2)	1 (2)	0
16. IE Favorites	18.2 (1)	3.8 (1.8)	3.2 (0.8)	3.8 (0)	2.8 (0)	2 (0)	0
18. IE New Window	7 (0.8)	3.8 (0.8)	2.2 (0)	1.6 (0)	1 (0)	1 (0)	0

a

# PeerPressure Performance

---

- Sick Machine Sensitivity Evaluation

Case	Machine 1	Machine 2	Machine 3
2	16 (0) / 1779	32 (2) / 1272	14 (0) / 1272
5	1 (0) / 182	1(0) / 566	1 (0) / 1657
6	1(0) / 2789	12(0) / 1777	12 (0) / 2017
14	1(0) / 105	1(0) / 84	1 (0) / 64
16	1 (0) / 302	2(0) / 3209	1 (3) / 1908

# Future Work and Discussion

---

- Multi-gene troubleshooting – Multiple sick entries among the suspects
- Cross-application misconfiguration
- Heavy customization of applications can break the assumption of strong conformance in most of the configuration entries
- GeneBank Maintenance – Privacy issue also involved

# Conclusion

---

- Automated diagnosis of misconfigurations is possible
- Statistical analysis works
- Ranking scheme follows power law distribution curve