DOI:10.1145/2909877

Rebecca T. Mercuri and Peter G. Neumann

Inside Risks
The Risks of Self-Auditing Systems

Unforeseen problems can result from the absence of impartial independent evaluations.

Over two decades ago, NIST Computer Systems Laboratory's Barbara Guttman and Edward Roback warned that "the essential difference between a self-audit and an external audit is objectivity."[6] In that writing, they were referring to internal reviews by system management staff, typically for purposes of risks assessment—potentially having inherent conflicts of interest, as there may be disincentives to reveal design flaws that could pose security risks. In this column, we raise attention to the *additional* risks posed by reliance on information produced by electronically self-auditing sub-components of computer-based systems. We are defining such self-auditing devices as being those that display internally generated data to an independent external observer, typically for purposes of ensuring conformity and/or compliance with particular range parameters or degrees of accuracy.

Our recent interest in this topic was sparked by the revelations regarding millions of Volkswagen vehicles whose emission systems had been internally designed and manufactured such that lower nitrogen dioxide levels would be produced and measured during the inspection-station testing (triggered by the use of the data port) than would occur in actual driving. In our earlier writings, we had similarly warned about voting machines potentially being set to detect election-day operations, such that the pre-election testing would show results consistent with practice ballot inputs, but the actual election-day ballots would not be tabulated accurately. These and other examples are described further in this column.

Issues

We are not suggesting that all self-auditing systems are inherently bad. Our focus is on the risks of explicit reliance *only* on internal auditing, *to the exclusion of any independent external oversight*. It is particularly where self-auditing systems have end-to-end autonomous checking or only human interaction with insiders, that unbiased external observation becomes unable to influence or detect flaws with the implementation and operations with respect to the desired and expected purposes.

Although many self-auditing systems suffer from a lack of sufficient transparency and external visibility to ensure trustworthiness, the expedience and the seeming authority of results can inspire false confidence. More generally, the notion of *self-regulation* poses the risk of degenerating into *no regulation whatsoever*, which appears to be the case with respect to self-auditing.

By auditing, we mean *systematic examination and verification of accounts, transaction records (logs), and other documentation, accompanied by physical inspection (as appropriate), by an independent entity*. In contrast, self-auditing results are typically internally generated, but are usually based on external inputs by users or other devices. The self-audited aggregated results typically lack a verifiable correspondence of the outputs with the inputs. As defined, such systems have no trustworthy independent checks-and-balances. Worse yet, the systems may be proprietary or covered by trade-secret protection that

explicitly precludes external inspection and validation.

Trade secrecy is often used to maintain certain intellectual property protections—in lieu of copyright and/or patent registration. It requires proofs of strict secrecy controls, which are inherently difficult to achieve in existing systems. Trade-secrecy protection can extend indefinitely, and is often used to conceal algorithms, processes, and software. It can thwart detection of illicit activity or intentional alteration of reported results.

Relying on internally generated audits creates numerous risks across a broad range of application areas, especially where end-to-end assurance is desired. In some cases, even internal audits are lacking altogether. The risks may include erroneous and compromised results, opportunities for serious misuse, as well as confusions between precision and accuracy.

Systemic Problems

Of course, the overall problems are much broader than just those relating to inadequate or inappropriately compromised internal auditing and the absence of external review.

Of considerable relevance to networked systems that should be trustworthy is a recent paper[2] that exposes serious security vulnerabilities resulting from composing implementations of apparently correctly specified components. In particular, the authors of that paper examine the client-side and server-side state diagrams of the Transport Layer Security (TLS) specification. The authors show that approximately a half-dozen different popular TLS implementations (including OpenSSL and the Java Secure Socket Extension JSSE) introduce unexpected security vulnerabilities, which arise as emergent properties resulting from the composition of the client-side and server-side software. This case is an example of an open source concept that failed to detect some fundamental flaws—despite supposed many-eyes review. Here, we are saying the self-auditing is the open source process itself. This research illustrates some of the risks of ad hoc composition, the underlying lack of predictability that can result, and the lack of auditing sufficient for correctness and security. However, their paper addresses only the tip of the iceberg when it comes to exploitable vulnerabilities of open source systems.

Digital Meters

The relative inaccuracy of self-calibrated (or merely factory-set) meters is often neglected in electronic measurement and design. Self-calibration can be considered to be a form of self-auditing when performed to a presumed reliable reference source. Calibration is also highly dependent on the specific applications. For example, while a 5% error rate may not be of tremendous concern when measuring a 5-volt source, at higher test levels the disparity can become problematic. There is also the error of perception that comes with digital displays, where precision may be misinterpreted as accuracy. Engineers have been shown to have a propensity toward overly trusting trailing digits in a numerical read-out, when actually analog meters can provide less-misleading relative estimates.[8]

Many concerns are raised as we become increasingly dependent on health-monitoring devices. For example, millions of diabetics test their blood glucose levels each day using computerized meters. System accuracy for such consumer-grade devices is recommended to be within 15 mg/dl as compared with laboratory results, yet experimental data shows that in the low-blood sugar range (<= 75 mg/dl), some 5% of these personal-use meters will fail to match the (presumably more stringent) laboratory tests. Reliance on results that show higher than actual values in the low range (where percentages are most critical) may result in the user's failure to take remedial action or seek emergency medical attention, as appropriate. Many users assume the meters are accurate, and are unaware

CREDIT TK

that periodic testing should be performed using a control solution (the hefty price of which is often not covered by health insurance). In actuality, since the control-solution test uses the same meter and is not a wholly independent comparison (for example, with respect to a laboratory test), it too may not provide sufficient reliability to establish confidence of accuracy.

End-to-End System Assurance

The security literature has long demonstrated that embedded testing mechanisms in electronic systems can be circumvented or designed to provide false validations of the presumed correctness of operations. Proper end-to-end system design (such as with respect to Common Criteria and other security-related standards) is intended to ferret out such problems and provide assurances that results are being accurately reported. Unfortunately, most systems are not constructed and evaluated against such potentially stringent methodologies.

Yet, even if such methods were applied, all of the security issues may not be resolved, as was concluded in a SANS Institute 2001 white paper.[1] The author notes that the Common Criteria "can only assist the IT security communities to have the assurance they need and may push the vendor and developer for [a] better security solution. IT security is a process, which requires the effort from every individual and management in every organization. It is not just managing the risk and managing the threat; it is the security processes of Assessment, Prevention, Detection and Response; it is a cycle." Rebecca Mercuri also points out[7] that certain requirements cannot be satisfied simultaneously (such as, a concurrent need for system integrity and user privacy along with assuredly correct auditability), whereas the standards fail to mitigate or even address such design conflicts.

The Volkswagen Case and Its Implications

Security professionals are well aware that the paths of least resistance (such as the opportunities and knowledge provided to insiders) often form the best avenues for system exploits. These truths were underscored when Volkswagen announced in September 2015

Relying on internally generated audits creates numerous risks across a broad range of application areas.

"that it would halt sales of cars in the U.S. equipped with the kind of diesel motors that had led regulators to accuse the German company of illegally [creating] software to evade standards for reducing smog."[5]

While Volkswagen's recall appeared at first to be voluntary, it had actually been prompted by investigations following a March 2014 Emissions Workshop (co-sponsored by the California Air Resources Board and the U.S. Environmental Protection Agency (EPA), among others). There, a West Virginia University research team working under contract for the International Council on Clean Transportation (ICCT, a European non-profit) provided results showing the self-tested data significantly underrepresented what occurred under actual driving conditions. These revelations eventually led to a substantial devaluation of Volkswagen stock prices and the resignations of the CEO and other top company officials, followed by additional firings and layoffs. Pending class-action and fraud lawsuits and fines promise to be costly in the U.S. and abroad.

Ironically, the report[9] was originally intended to support the adoption of the presumably strict U.S. emissions testing program by European regulators, in order to further reduce the release of nitrogen oxides into the air. Since the university researchers did not just confine themselves to automated testing, but actually drove the vehicles on-road, they were able to expose anomalous results that were as much as 40 times what is allowed by the U.S. standard defined by the Clean Air Act. The EPA subsequently recalled seven vehicle models dating from 2009–2015, including approximately 500,000 vehicles in the U.S.; Germany ordered recall of 2.4M vehicles. Extensive hardware and software changes are required to effect the recall modifications. Still, the negative environmental impacts will not be fully abated, as the recalls are anticipated to result in poorer gas mileage for the existing Volkswagen diesel vehicles.

Election Integrity

An application area that is particularly rife with risks involves Direct Recording Electronic (DRE) voting systems—which are self-auditing. These are end-to-end automated systems, with results based supposedly entirely on users' ballot entries. Aggregated results over multiple voters may not have assured correspondence with the inputs. Most of the commercial systems today lack independent checks and balances, and are typically proprietary and prohibited from external validation.

Reports of voters choosing one candidate and seeing their selection displayed incorrectly have been observed since the mid-1990s. This occurs on various electronic balloting systems (touchscreen or push-button). However, what happens when votes are recorded internally (or in processing optically scanned paper ballots) inherently lacks any independent validation. For example, Pennsylvania certified a system even after videotaping a vote-flipping incident during the state's public testing. The questionable design and development processes of these systems—as well as inadequate maintenance and operational set-up—are known to result in improper and unchecked screen alignment and strangely anomalous results.

Some research has been devoted to end-to-end cryptographic verification that would allow voters to demonstrate their choices were correctly recorded and accurately counted.[4] However, this concept (as with Internet voting) enables possibilities of vote buying and selling. It also raises serious issues of the correctness of cryptographic algorithms and their implementation, including resistance to compromise of the hardware and software in which the cryptography would be embedded.

Analogous Examples

It seems immediately obvious that the ability to rig a system so it behaves cor-

rectly *only* when being tested has direct bearing on election systems. The Volkswagen situation is a bit more sophisticated because the emissions system was actually controlled differently to produce appropriate readings whenever testing was detected. Otherwise, it is rather similar to the voting scenario, where the vendors (and election officials) want people to believe the automated testing actually validates how the equipment is operating during regular operations, thus seemingly providing some assurance of correctness. While activation of the Volkswagen stealth cheat relied on a physical connection to the testing system, one might imagine a tie-in to the known locations of emission inspection stations—using the vehicle's GPS system—which could similarly be applied to voting machines detecting their polling place.

Election integrity proponents often point to the fact that lottery tickets are printed out by the billions each year, while voting-system vendors seem to have difficulty printing out paper ballots that can be reviewed and deposited by the voter in order to establish a paper audit trail. Numerous security features on the lottery tickets are intended to enable auditing and thwart fraud, and are in principle rather sophisticated. While the location and time of lottery ticket purchases is known and recorded, this would not be possible for elections, as it violates the secrecy of the ballot. However, it should be noted that insider lottery fraud is still possible, and has been detected.

Automatic Teller Machines (ATMs) are internally self-auditing, but this is done very carefully—with extensive cross-checking for consistency to ensure each transaction is correctly processed and there are no discrepancies involving cash. There is an exhaustive audit trail. Yet, there are still risks. For example, some ATMs have been known to crash and return the screen to the operating-system command level. Even more riskful is the possible presence of insider misuse and/or malware. Code has been discovered for a piece of malware that targets Diebold ATMs (this manufacturer was also a legacy purveyor of voting machines). The code for this malware used undocumented features to create a virtual 'skimmer' capable of recording card details and personal identification numbers without the user's knowledge, suggesting the creator may have had access to the source code for the ATM. While this does not directly point to an inside job, the possibility certainly cannot be ruled out. Experts at Sophos (a firewall company) believe this code was intended to be preinstalled by an insider at the factory, and would hold transaction details until a special card was entered into the machine—at which point a list of card numbers, PINs, and balances would be printed out for the ne'er-do-well to peruse, and perhaps use, at leisure. It is also possible the malware could be installed by someone with access to the ATM's internal workings, such as the person who refills the supply of money each day (especially if that malware were to disable or alter the audit process).

Complex Multi-Organizational Systems

One case in which oversight was supposedly provided by corporate approval processes was the disastrous collapse of the Deepwater Horizon. The extraction process in the Gulf of Mexico involved numerous contractors and subcontractors, and all sorts of largely self-imposed monitoring and presumed safety measures. However, as things began to go wrong incrementally, oversight became increasingly complicated—exacerbated further by pressures of contractual time limits and remote managers. This situation is examined in amazing detail in a recent book on this subject.[3]

Conclusion

Recognition of the risks of systems that are exclusively self-auditing is not new. Although remediations have been repeatedly suggested, the reality is even worse today. We have a much greater dependence on computer- and network-based systems (most of which are riddled with security flaws, potentially subject to external attacks, insider misuse, and denials of service). The technology has not improved with respect to trustworthiness, and the total-system risks have evidently increased significantly.

Independent verification is essential on a spot-check and routine basis. Security must be designed in, not added on; yet, as we have seen, hacks and exploits can be designed in as well. Hired testers may suffer from tunnel vision based on product objectives or other pressures. Group mentality or fraudulent intent may encourage cover-up of detected failure modes. Whistle-blowers attempting to overcome inadequate self-auditing are often squelched—which tends to suppress reporting. Classified and trade secret systems inherently add to the lack of external oversight.

The bottom line is this: Lacking the ability to independently examine source code (much less recompile it), validate results, and perform spot-checks on deployed devices and system implementations, various anomalies (whether deliberate or unintentional) are very likely to be able to evade detection. Specific questions must be periodically asked and answered, such as: What independent audits are being performed in order to ensure correctness and trustworthiness? When are these audits done? Who is responsible for conducting these audits? Without sufficient and appropriate assurances, self-auditing systems may be nothing more than a charade. C

References

1. Aizuddin, A. The Common Criteria ISO/IEC 15408—The Insight, Some Thoughts, Questions and Issues, 2001; http://bit.ly/1IVwAr8
2. Beurdouche, B. et al. A messy state of the union: Taming the composite state machines of TLS. In *Proceedings of the 36th IEEE Symposium on Security and Privacy*, San Jose, CA (May 18–20, 2015); https://www.smacktls.com/smack.pdf
3. Boebert, E. and Blossom, J. *Deepwater Horizon: A Systems Analysis of the Macondo Disaster.* Harvard University Press, 2016.
4. Chaum, D. Secret-ballot receipts: True voter-verifiable elections. *IEEE Security and Privacy 2*, 1 (Jan./Feb. 2004).
5. Ewing, J. and Davenport, C. Volkswagen to stop sales of diesel cars involved in recall. *The New York Times* (Sept. 20, 2015).
6. Guttman, B. and Roback, E.A. *An Introduction to Computer Security: The NIST Handbook.* U.S. Department of Commerce, NIST Special Publication 800-12 (Oct. 1995).
7. Mercuri, R. Uncommon criteria. *Commun. ACM 45*, 1 (Jan. 2002).
8. Rako, P. What's all this meter accuracy stuff, anyhow? *Electronic Design 16*, 41 (Sept. 3, 2013).
9. Thompson, G. et al. In-use emissions testing of light-duty diesel vehicles in the United States. International Council on Clean Transportation (May 30, 2014); http://www.theicct.org

Rebecca Mercuri (notable@notablesoftware.com) is a digital forensics and computer security expert who testifies and consults on casework and product certifications.

Peter G. Neumann (neumann@csl.sri.com) is Senior Principal Scientist in the Computer Science Lab at SRI International, and moderator of the ACM Risks Forum.