Security Metric of the Week #16:
Number of security policy noncompliance infractions detected

Authors:
Krag Brotby, CISM, CGEIT
Gary Hinson, PhD, MBA, CISSP

The extent to which employees comply with the organization's security policies sounds like the kind of thing that management might want to track and, where appropriate, improve. This week's metric is a typical, if rather naive attempt to measure policy compliance ... by counting noncompliance incidents.

Policies are 'mandated', in other words management expects everyone to comply with them unless there are justified reasons not to comply (meaning authorized exemptions for those organizations that are mature enough to appreciate the need to manage this aspect carefully). While management originates most of the security requirements documented in policies, some derive from external obligations under applicable laws, regulations or agreements with third parties (e.g. PCI-DSS).

The metric's wording implies that unauthorized noncompliance 'infractions' (more often called incidents) are being detected and recorded in a form that can be counted - typically some sort of security incident database, usually part of a Help Desk ticket management system. Why not simply report the number of security incidents recorded by the Help Desk over, say, a month? Such a metric would be low cost, but what about its benefits?

In reality, many other noncompliance situations occur, and some of them are detected or identified but for various reasons don't get reported as incidents. As an example, who would bother reporting an everyday tailgating incident, even in a military organization that prides itself on physical security? Furthermore, lots more noncompliance incidents are not even identified as such - they are not observed or recognized, or they are very short-lived or otherwise deemed trivial. If an employee spots and challenges a tailgater, who then proceeds to identify themselves with an authentic-looking staff pass, the 'incident' is such a non-event that it is most unlikely to be reported, but it could of course be an actual intrusion.

All of this constitutes a huge bias to the metric as worded, a tremendous source of random error or noise in the measurement values.

Maybe it would help if we clarified the metric by reporting not the absolute number of policy noncompliance incidents but the rate of occurrence i.e. the number of incidents in a predefined period. What would it mean if the metric

jumped from 97 to 145 one month? Is that something that should concern management? What if it went from 145 to 97, or from 97 to 99, or hit zero? How, exactly, would this metric support the decision making process? Precisely what would management be expected to do with it?

Probing questions of this nature soon belie this metric's superficial allure. It is not hard to find fault with it. Arguably the most fundamental issue is that it is practically impossible to determine the true number of noncompliance incidents by direct observation, except perhaps in strictly controlled experimental conditions. The best we can reasonably hope achieve in reality is to estimate the true number as rationally and accurately as we can, for instance by statistical means, using a more scientific process for identifying and reporting noncompliance incidents, such as periodic security policy compliance audits. Unfortunately, that approach substantially drives up the Cost of the metric and so adversely affects its PRAGMATIC score:

| P | R | A | G | M | A | T | I | C | Score |
|---|---|---|---|---|---|---|---|---|---|
| 55 | 64 | 75 | 50 | 68 | 34 | 59 | 76 | 33 | 57% |

We have been quite generous on the 75% rating for Actionability on the assumption that, if the measurements were poor, management would initiate whatever they considered appropriate to improve policy compliance, such as training and awareness activities coupled with increased management oversight, and perhaps more emphasis on enforcement actions. We didn't have the same latitude with the rating for Accuracy, although using auditors or other professional assessors to measure the metric could improve its Independence, relative to self-reporting of noncompliance incidents by employees.

The Genuineness rating suffers largely because, if this metric were being reported and used proactively by management in an attempt to improve policy compliance, there is a distinct possibility that it would be deliberately manipulated. There are some obvious if crude ways in which employees might 'game the system' to drive up the metric without materially improving compliance, such as consciously failing to report noncompliance incidents. Even if management succeeded in addressing these tricks (e.g. by instituting and enforcing a policy on reporting incidents), other more subtle games would probably flourish. It is amazing how creative people can get in the face of adversity!

The Predictability rating is also depressed because it is a backwards-looking metric: it tells us how things were in the preceding period but doesn't say

much about how they may change in the forthcoming period, other than vague indications that might emerge from the background noise.

It is a reasonable assumption that security policies themselves are directly Relevant to information security, hence compliance with the policies is also Relevant. However, there is more to security than policy compliance (it is 'necessary but not sufficient'), and as noted elsewhere the metric as worded does not perfectly reflect policy compliance, hence we rated the metric 64% on the Relevance criterion. [This paragraph illustrates the PRAGMATIC thinking behind each of the ratings. If we had the time and energy, we should probably document all the ratings on all the metrics for future reference, but in practice we would be more inclined to elaborate on and write-up the rationales for the few security metrics that we eventually adopt.]

The overall PRAGMATIC score of 57% tells us that, as originally stated, this is unlikely to feature as one of the few good security metrics we would chose, unless perhaps we were so short of inspiration that there were no higher-scoring candidate metrics on the table.

Having discussed our concerns and hinted at some of the ways in which we might improve this metric, do you have some even better suggestions? Or do you agree that this metric is essentially doomed? Submit a comment and we'll do our best to respond positively.