USING CHECKLISTS TO MAKE BETTER BEST

Craig S Wright and Tanveer A Zia
School of Computing and Mathematics
Charles Sturt University
Wagga Wagga, NSW, Australia
{crwright,tzia}@csu.edu.au

Abstract

The more routine a task is we see the greater the need for a checklist. Even the smartest of us can forget where we parked our cars on returning from a long flight. So, the question is, why not create a straightforward checklist that will improve system management and security? In Information Technology operations, the vast majority of skilled people have re-built servers, but in an incident response situation, it can be unforgivable to overlook a serious security configuration simply because in the stress of the environment causes one to lose track of which stage they were on while being interrupted and multitasking. We show that the use of standard checklists and flowcharts created by the individual make for better results even in daily tasks. This paper presents the results of an experiment into the use of checklists by incident responders. It demonstrates how basic checklists can improve an organisation's security.

Keywords
Security, Checklists, Incident handling, Incident response

INTRODUCTION

It is reasonable to conclude that as we improve and increase the level of skills and proficiency in a task, we no longer need to refer to checklists and reminders for help. This comes to a belief that somehow we will remember everything and that as seasoned professionals, we no longer need to be impelled to remember the steps in a complex process. In many instances, this may be true, but the outlier remains in high stress situations as well as when tasks are no longer a daily chore. This paper provides research that demonstrates the discrimination against using checklists in the belief that it is harmful to incident response. It also shows that the creation of a simple checklist of steps by an incident responder before an event will minimise the number of errors and outlier events in incident response.

This study was started in 2009. The work of Gawande (2009) was reviewed by the author after the release of the preliminary results from this research. Gawande showed that the hubris and clear distain of checklists remains a common experience across multiple professions (although not all). In this work, a number of examples and studies were cited demonstrating how the use of a straightforward checklist could improve the results of common medical procedures and save lives.

Yet, just as in Information Technology, physicians and others in highly specialised professions shun the use of a basic checklist that has been shown to increase patient safety. This paper presents research demonstrating how the use of a basic checklist reduces the number of false positives. These checklists have been created by the responders using their own processes and steps.

LITERATURE REVIEW AND RELATED WORK

Considering the wide range of security papers calling for the use of checklists (Baskerville, 1993; Martin, 1973; Dhilon & Backhouse, 2001; Lincoln, 1998), remarkably little research is available quantifying the effects of using checklists to reduce the risk a system is exposed to. Checklists have evolved over time and proponents of lists (Elberzhager, Klaus, & Jawurek, 2009) have created checklists for about every conceivable situation.

Bishop and Frincke (2005) tell us that "security checklists cause both alarm (because they can replace thinking with conformance to irrelevant guidelines) and delight (because they serve as aids in checking that security considerations were properly taken into account)" and use both analogy and anecdote to demonstrate that checklists can boost security testing.

Bellovin (2008) reminds us that a poorly structured checklist "especially if followed slavishly or enforced without thought—can make matters worse." But little is also provided as to what effect a checklist can include.

In this study, the individuals have been allowed to create their own checklist, both to minimise any potential aversion to using such a tool as well as to align this to the existing best practices of the individual and organization.

METHODOLOGY

The experiment was created as a simple analysis in order to minimise any impact on existing operations in firms that allowed this research to happen to utilise their people and operations. The names of the firms have been withheld due to operational considerations and the requirement that the names of the organisation remain confidential.

The experiment involved the analysis of incident response personnel as they reacted to incidents. The study was based on the response times with and without a checklist. The same individuals were monitored and the response times were measured with the use of a checklist and without. These are ordinary people in the course of their daily activities. When the test was started, none of the individuals used a checklist. All of the individuals are highly skilled and maintain industry certifications such as GCIA and GCIH related to their roles. The manager in the leading organisation was 6-Sigma trained and was receptive to applied experimentation of this nature.

The checklist requirement stated that the individual had to create a checklist of at least one and at most two pages in length. A flow diagram was allowed. The analysts were asked to create their own checklist using their own processes. These are the steps the responder would expect in any normal incident. The experiment required that the analyst had to get out and read the checklist at the start of the incident when the checklist was used. If any significant event occurred during the incident process, the checklist was to be read again.

The results were measured based on the time taken to respond. A later analysis of the results was conducted based on the recorded data at the organisation which was used to measure error rates in an effort to improve responder capabilities.

It was expected that the results themself would change as people encounter good and bad days at work.. To account for this variability and as diverse incidents would appear on any given day, the responder would toss a virtual coin. This was a small process loaded into the web portal. The system was set to change to show a green icon on the analysts' screen in the event that they were to use a checklist and to present as blue when they were not to use the checklist.

To define the variables and hypothesis.

The hypothesis was formulated prior to the start of the experiment and can be stated that there would be no statistically significant difference in mean results between the times taken from the start of an incident or event to the determination that an event had or had not occurred. This is the value we measured and was defined as time in minutes "*t"*.

It could even state that if "*t*" was larger for those with a checklist, the result was negative and a checklist made things worse. If it was found that "t" was smaller for those incidents with a checklist than those with one, we could say that a checklist improved the incident response process. In this test, the responders were all highly skilled. It would be expected that any positive results found for a highly trained responder would be more beneficial for an inexperienced security professional or even a more generalised IT professional (Bishop & Frincke 2005).
The experiment measured the following variables:

- t_{ij} Here i is the i^{th} individual with the value 'j' in minutes to establish a response and determine if an event was an incident or not.
- $t_{ij(check)}$ This is the subset of readings where the individual 'I' used a checklist as measured in minutes
- $t_{ij(free)}$ This represents the subset of readings where the individual 'I' did not use a checklist.

Using these variables allowed for the calculation of the following:

- $t_{i(ave)}$ This is the average response time in minutes for an individual 'I'.
- $t_{i(check)}$ This is the average time for the individual to respond and determine if an event is an incident using the checklist.
- $t_{i(free)}$ This is the average time for the individual to respond and determine if an event is an incident without using the checklist.

With these values, the study and hypothesis can be defined particularly well.
First, define Ho as the null hypothesis and Ha as the alternative hypothesis. Next, it is necessary to express our hypothesis as follows:

Ho $t_{i(check)} = t_{i(free)}$
Ha $t_{i(check)} <> t_{i(free)}$

The null hypothesis is that there is no difference in how long it will take an individual on average using a checklist to respond and establish if an event is an incident or not. The alternative hypothesis is that the use of a checklist will result in a difference in how long the responder reacts. This is testing if the time taken to respond using a checklist will be significantly different to that without a checklist.

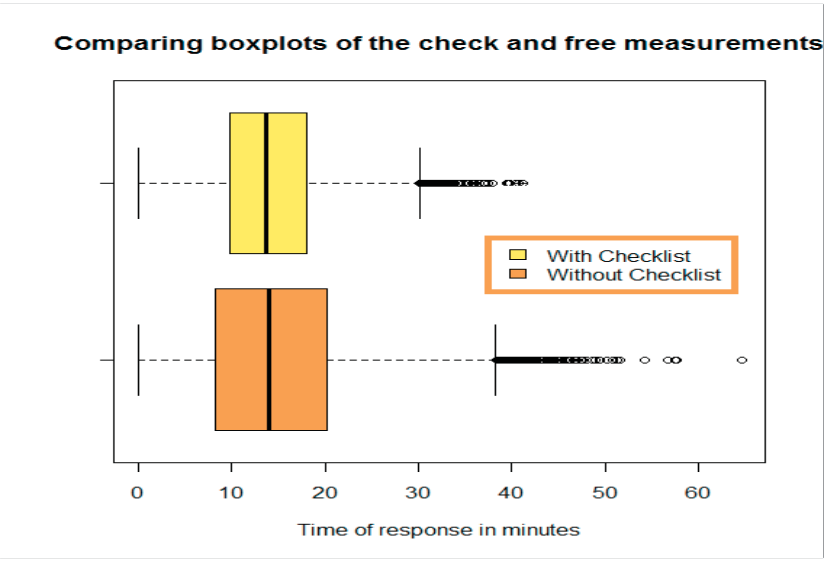

Figure 1: Comparing response times with and without checklists

Although each event will vary in style and structure and the responder will vary in capacity throughout the day and at different points in their lives, the averages when taken over time should be the same (United States Fire Administration, 2004). In order to ensure that the responders would follow the process, the responders created and used their own checklists based on the best practice as they determined and defined it. These controls removed arguments over best practice and individualised processes that would have been expected to occur.

The responder bias was minimised using a system of randomising the times when a responder would use the checklist. The study did not just set times of the day used in the process. Also, effects from time of day and week were also removed. As a virtual coin toss determined if the responder used the checklist or not, the responder did not select the responses they would choose to use a checklist on or not. There are limitations to this, but we all have to work within the constraints of the world and empirical studies on active companies and with existing incidents need to be measured in a manner that allows the organisation to function as it is being experimented on.

RESULTS

The results of the study are presented in Figure 1 as a boxplot. In simply visualising the two datasets, it is possible to determine that there is a difference in the standard deviations with a larger range of values for the responses without a checklist than those recorded when a checklist was used. Looking at the statistics in R (our statistical package and displayed in Table 1), we see a mean (average) value of 14.3602 minutes for responses without the use of a checklist and 14.00188 when a checklist is used.

	Minimum	1st Quartile	Median	Mean	3rd Quartile	Max
ticheck	0.000	9.878	13.680	14.000	18.000	41.260
tifree	0.000	8.246	14.060	14.360	20.280	64.600

Table 1: A comparison of ticheck (time with a checklist) against tifree (the time without using a checklist).

The mean values are only 21 seconds different on average over a mean of around 14 minutes. This visual inspection does not show the complete result. By conducting a Student's t-test on the two datasets, we can see if a difference in the values actually exists or not. This is easy to do in R and the results are displayed below (as output from R).

```
> t.test(tifree, ticheck)

        Welch Two Sample t-test

data:  tifree and ticheck
t = 3.3964, df = 20638.23, p-value = 0.000684
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1515310  0.5651013
sample estimates:
mean of x mean of y
 14.36020  14.00188
>
```

What this all means is that at the alpha = 5% level we have a p-value of 0.000684 and we are confident that there is a statistically significant difference in the means. The results indicate that we can confidently reject the Null hypothesis and accept the alternative hypothesis Ha that a checklist does make a positive effect on the diagnosis of an incident.

Type I error and monitoring intrusions

The results incident analysis and the rate of error in diagnosis was measured using an experiment where noted incidents where replayed. Existing PCAP capture traces from sites with known attack and incident patterns were loaded into an analysis system for evaluation purposes. The OSSIM and BASE frontends to snort had been deployed for this exercise. SQL scripts where altered to display a random lag into the responses and tcpdump was used to replay the PCAP trace as if it occurred *'live'*. The analyst had to decide if each incident was worth escalating or should be noted and bypassed. The results of this exercise are reported in Figure 2 through a display of type I errors.

It is easy to see from Figure 1 that as the response time of the system increases, so does the analyst's error rate. The lag in returning information to the analyst has a direct causal effect. We see that the longer the lag between requesting the page and that where the page is returned, the greater the error rate in classifying events.

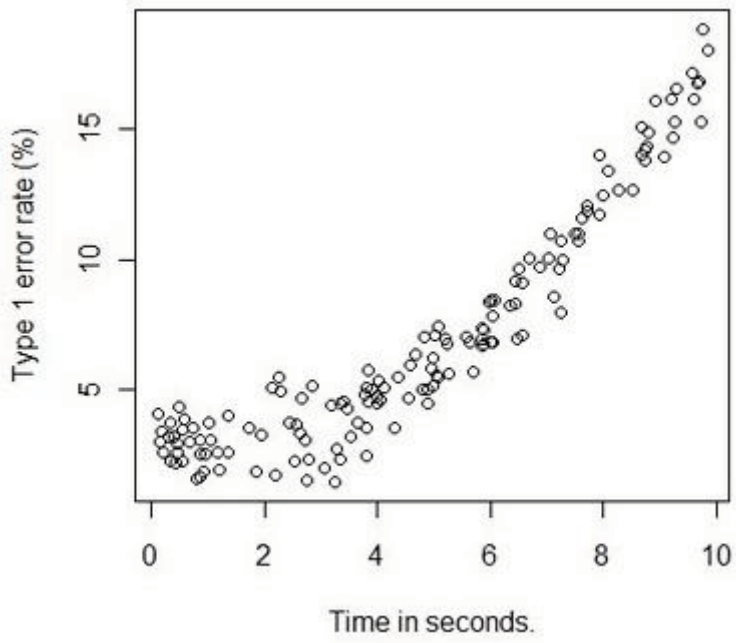

Figure 2: Type I error rate and the time to respond.

The analysis of this data was extended to include a Loess calculated plot of the expected error rate against time (Figure 2).

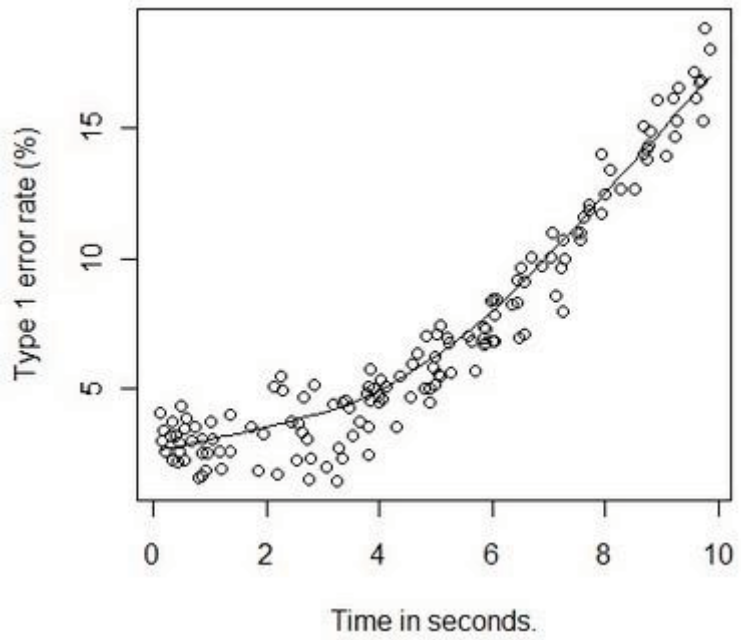

Figure 3: A Loess plot of the Type I error rate for the time to respond.

Using this plot, we can clearly see that the slope increases sharply after around 4 seconds. As such, it is essential to ensure that responses to queries are returned in under 3-4 seconds. From this, we understand that time delays significantly affect the performance of an incident handler.

DISCUSSION

Although there is little difference in the mean value returned from the use of a checklist to that when the responder did not use a checklist, there are some outliers where something has gone wrong. We see from the boxplot (Figure 1), that there are occasions without a checklist where errors do exist. These mistakes increased the time required for the incident response process. In particular, there is a long tail effect when no checklist has been used.

In responding to an incident, having a checklist helps even experienced professional incident responders. A professional may have worked in the same capacity for many years and understand his work implicitly, yet, under stress or in a rush, there may be things missed or overlooked.(Hales & Pronovost 2006).

The null hypothesis can be stated as there is no difference in how long it will take an individual on average to respond and to determine whether an event is an incident or not when we measure the time taken using a checklist against not using one. The alternative hypothesis is that the use of a checklist will result in a difference in how long the responder reacts. We can state that the time measured using a checklist will be significantly different to that without a checklist.

When the initial result is coupled with the result (section 3.1) that time delays increase Type I errors, it is clear that using a checklist is an effective low cost solution to improving the security and response times against an attack.

Each IDS system has an expected TYPE I and TYPE II error rate that will change as the system is tuned to the organisation's operating environment and with the input of the responder. The result of this is that this is a peculiar function for the organisation that can only be approximated for other organisation (even when the same IDS product is deployed). The inherent accuracy of the IDS (which is a trade-off between TYPE I and TYPE II errors and it is a cost function in itself) will be dependent on the input of the individual assessing it. The IDS forms a cost function as the increase in reporting results in a greater number of false positives that need to be investigated. In limiting the false positives, the likelihood of missing an event of note also increases. Each

validation of a false positive takes time and requires interaction from an analyst. Hence the tuning of an IDS is balanced on maximizing the detection rate against cost.

We see that adding reasonable controls that reduce error have value and improve the security of a site.

CONCLUSION

Using a checklist is beneficial in an incident response.. An exceptional incident responder is not afraid to use a checklist and develop a process. In fact, a decent responder will be a better responder directly through the creation of a two page checklist and/or flow diagram.

It is easy to see that the more routine a job is, the greater the need for a checklist (Turner, 2001). Even the smartest of us can forget where we parked our cars on returning from a long flight (Mann 2002). So we should be asking why not make a basic checklist that will make things better? In Information Technology operations, the vast majority of knowledgeable people have re-built servers, but in an incident response environment, it can be unforgivable to overlook a serious security configuration simply because the stress of the environment can cause one to lose track of which stage they were on while being interrupted and multitasking (SANS 2011). We show that the use of straightforward checklists and flowcharts created by the individual make for better results even in daily tasks.

Eliminating error can be directly correlated to the process of investigation and retesting in order to prevent future errors. In place of the negative incentives where individuals are punished for their failures, we should make a system that rewards exemplary practice. We can start this by encouraging security professionals to use checklists. Instead of expending effort to assigning individual responsibility and then punishing that individual, we should improve the process and practice (Bellovin, 2009).

A checklist will enhance this process. A checklist also allows us to determine the processes that function better and to integrate these into the system improving practice over time.

No one is infallible (Gawande, 2007) and "*to err is human*" (Cicero). However, by using checklists, you can at least be sure that your crises happen over the hard things, and not the easy things. This is particularly relevant to incident handling.

REFERENCES

Baskerville, R. (1993) "Information Systems Security Design Methods: Implications for Information Systems Development" ACM Computing Surveys, 25 (4), 375-414.

Bellovin, S. (2008), "Security by Checklist," Security & Privacy, IEEE , vol.6, no.2, pp.88, March-April 2008

Bellovin, S. (2009), "Security Analysis I" COMS W4187 — Security Architecture and Engineering (Fall '09) Columbia USA

Bishop, M.; Frincke, D.A.; (2005) , "Teaching secure programming," Security & Privacy, IEEE , vol.3, no.5, pp. 54- 56, Sept.-Oct. 2005

Dhillon, G. & Backhouse, J. (2001), "Current directions in IS security research: towards socio-organizational perspectives". Information Systems Journal, 11: 127–153.

Elberzhager, F., Klaus, A., & Jawurek, M. (2009) "Software Inspections Using Guided Checklists to Ensure Security Goals," Availability, Reliability and Security, International Conference on, pp. 853-858, 2009 International Conference on Availability, Reliability and Security.

Gawande, A. (2007) "The Checklist", Annals of Medicine, The New Yorker.

Gawande, A. (2009) "The Checklist Manifesto: How to Get Things Right" Macmillan, USA

Hales, B., & Pronovost, P. (2006) "The checklist—a tool for error management and performance improvement, Journal of Critical Care", Volume 21, Issue 3, September 2006, Pages 231-235, ISSN 0883-9441, 10.1016/j.jcrc.2006.06.002.

Mann, C., (2002) "Homeland Insecurity" The Atlantic Monthly

Martin, J. (1973) "Security, Accuracy and Privacy in Computer Systems", Prentice Hall USA

SANS (2011) "Hacker Techniques, Exploits & Incident Handling" SANS Institute, USA

Turner, T., (2001) “Controlling Pilot Error: Checklists and compliance”, McGraw Hill Professional USA

United States Fire Administration, Federal Emergency Management Agency. (2004). “Responding to Incidents of National Consequence: Recommendations for America’s Fire and Emergency Services Based on the Events of September 11, 2001, and Other Similar Incidents”. (FA-282-May 2004).