A QUANTITATIVE EXAMINATION OF THE CURRENT STATE OF CORPORATE SECURITY PRACTICES

Clint Gibler
NCC Group Domain Services, USA

Email clint.gibler@nccgroup.com

ABSTRACT

In order to augment and scale limited in-house security expertise, many organizations rely on automated security scanning tools to find misconfigurations, services that need to be patched, and web application vulnerabilities. While much research has been done into detecting new types of vulnerabilities and finding known ones more precisely, there has been disappointingly little examination of how successful these techniques are in practice and, more importantly, how effective these tools are in making companies more secure.

We will discuss insights gained from analysing the results of running a commercial security scanner on 100 international companies across 10 industry verticals from February 2014 to May 2015, collectively representing over 900,000 findings. We examine questions such as: what are the common types of vulnerabilities in real companies today? Does it vary by industry? For a given type of vulnerability, how long does it take companies to remediate issues? Does the time to fix depend on one or more of: the type of the vulnerability, its severity, or merely on its solution? Do companies or industries tend to fix the same types of vulnerabilities in a similar time frame or is there significant variation?

We aim to provide industry professionals with objective data against which they can compare their company's performance, and security researchers with insights into impactful areas they can focus on in their future work.

INTRODUCTION

Technology continues to advance at a rapid rate, and with increased focus from determined attackers, organizations are feeling more pressure than ever to keep their websites and infrastructure secure. Despite this need, there is a severe shortage in the security professional labour market; in fact, some predict a 1.5 million job difference between the security professional workforce demand and talent available over the next five years [1]. The US government, companies and universities are working to address this shortage by holding Capture The Flag (CTF) competitions and other challenges to attract talent, such as NetWars [2] or the US Cyber Challenge [3], as well as adding security courses to undergraduate curriculums and providing developers with security training.

All of these approaches are important steps towards helping address the security skills shortage. However, in the near term, security professionals are likely to continue to be outnumbered five or more to one by developers in their organization. To scale their efforts, security professionals often create their own tools or leverage existing automated security scanning tools to find misconfigurations, services that need to be patched, and web application vulnerabilities.

Over time, security professionals become familiar with their company's software development lifecycle (SDLC) and the types of security issues their projects tend to face. However, it is difficult to gain insight into the types of security issues that plague one's industry or companies in general, as companies are generally reluctant to discuss any lapses in security publicly due to potential negative PR. While this reticence is expected, it makes it difficult for security practitioners and researchers to focus on the more prevalent and thus impactful areas of security facing companies today.

In this paper, we are in the fortunate position of being able to discuss vulnerability trends across many companies. Our dataset consists of over 908,000 findings from running a commercial security scanning tool on 100 companies in 10 industries from February 2014 to May 2015. We analyse in detail topics including:

- The common types of vulnerabilities across all companies and by industry.
- The *time to fix* of vulnerabilities – how effectively do various industries handle different categories of vulnerabilities? Is the time to fix of a vulnerability affected by the severity (as approximated by CVSS score) or its remediation solution?

The results of our analysis can provide industry professionals with objective data against which they can compare their company's performance and security researchers with insights into impactful areas on which they can focus in their future work.

Key takeaways

We observe the following key takeaways from our evaluation:

- There can be significant value to companies in using a managed service that vets tool findings and removes false positives, as the false positive rate for industries ranges from 49%–89%. At this rate, a company would need to invest nontrivial resources to extract actionable results from the scans.
- Across all companies, the most prevalent findings belong to the following categories: issues related to the server or language running a web application, PKI/SSH/SSL, web application best practices and general network issues.
- Of the findings that end up being resolved, a large percentage of them are addressed within the first 10–20 weeks after discovery. While the rate of addressing findings tends to taper after 30 weeks, surprisingly, some findings continue to be resolved past 50 weeks.
- There doesn't appear to be a strong correlation between CVSS score and time to fix, though higher CVSS findings tend to be resolved at a higher rate.
- Of the high level categories, host findings were resolved at the highest rate, followed by web application and lastly network findings. There's a large disparity between industries in the percentage of web application and network findings resolved: up to 20%–40%.
- Regarding time to fix by remediation solution:
  - In general, a high percentage of findings that require updating a language or package tend to be resolved, usually within 20 weeks.

- Findings that require more environment-specific solutions, such as disabling a vulnerable service, reconfiguring default accounts, removing files, and updating configuration, tend to take longer to be resolved and are more likely to be left unresolved.
- Cryptography-related issues, such as those requiring reconfiguring SSL or updating OpenSSL or SSH, have the lowest likelihood of being resolved, ranging from about 70%–80%.

DATASET

Our dataset consists of slightly over 908,000 findings that are the result of running a security scanning tool on 100 companies across 10 industries from February 2014 to May 2015. Companies were randomly selected within industries and were scanned from a handful of times to over 100 times. Most industries are represented by ten companies, with the exception of the energy & utilities and health industries, with six and nine companies respectively, and the financial services industry with 15.

Each scan contains information about when it was started, the tool used and the mode in which the tool was run, a GUID representing the asset group scanned, and a list of findings. Each finding consists of a URL or IP address, a port, a descriptive title, a CVSS v2 score, an optional set of remediation solutions, and whether the finding is a true positive or a false positive (TP or FP).

All findings were vetted manually by a Technical Account Manager (TAM) and deemed to be a true or false positive. However, some findings were labelled as false positives that were in fact true positives because the tool reported many duplicate findings and the TAM wished to reduce the noise presented to the customer, as there was no state for 'ignore' in the system at the time. This affects the true/false positive rates we report, but since all of the trends and analyses in the evaluation section are based only on true positive findings, this inaccuracy causes our figures and results to be in the worst case underestimates of reality.

Though several groups have voiced concerns about the accuracy of CVSS scores [4, 5], as our dataset is already labelled with CVSS v2 scores we use them as an approximation for severity in this work. We leave leveraging more precise severity metrics to future work.

We note that our dataset is quite heterogeneous – there is a wide range between the amount of data available both between companies and between industries. This variety causes difficulty when one tries to make overall statements or generalizations. Ideally, each company would be scanned an equal number of times, by the same mode of the same tool, and in the same time increments. As real-world datasets are never ideal, we have attempted to be cautious in the conclusions we draw in this paper. Further challenges caused by the heterogeneity of the data are described in the 'Time to fix' subsection of the 'Methodology' section.

METHODOLOGY

In this section we describe the categories into which we group our findings as well as how we calculate a finding's time to fix, both of which are used extensively in the figures in the 'Evaluation' section.

Categories

As our dataset contains thousands of unique types of vulnerabilities, we group findings into three high-level categories:

- **Host issues** – host-based vulnerabilities such as those related to the OS, a running database, or various out-of-date software.
- **Network issues** – vulnerabilities related to network communication, such as those involving PKI, SSH or SSL, credentials for a service being transmitted in cleartext, insecure remote services, etc.
- **Web application issues** – vulnerabilities in the web server or language in which the web application is written, failure to meet web application best practices or common web application vulnerabilities like cross-site scripting (XSS) or cross-site request forgery (CSRF).

We further break down each of these three category groups into three to six more specific categories so that more detailed insights may be drawn. The categories are largely self-explanatory but we provide additional descriptions and example types of vulnerabilities in each category in Appendix A.

Time to fix

We wish to gain insight into the time it takes companies and

Industry	# Companies	Total # scans	Avg scans	# Findings	% TP
Charities	10	293	29.3	42,184	51%
Energy & utilities	6	28	4.67	31,391	11%
Financial services	15	339	22.6	120,158	48%
Health	9	23	2.56	14,220	17%
IT	10	246	24.6	84,803	42%
Leisure & media	10	149	14.9	257,685	16%
Public sector – education	10	101	10.1	155,499	23%
Public sector – local	10	95	9.5	124,363	23%
Retail	10	101	10.1	44,314	50%
Transport	10	318	31.8	33,483	23%

Table 1: Composition of dataset.

industries to address different types of findings, to see if it varies by company, industry, or by some property of the finding itself.

There are several subtle aspects that must be taken into account when calculating a finding's time to fix. First, what uniquely identifies a finding? An initial approach might say that a finding can be characterized by the specific vulnerability detected (for example, a particular *Apache* CVE), the URL or IP on which it was detected, and the port. However, with the increased prevalence of cloud hosting such as *Amazon EC2*, the same IP may belong to different companies at different points in time. Thus, considering both the scan time and company is important.

When is a finding *resolved*? At a high level, a finding is resolved when the same company is scanned again and the finding is no longer detected. However, a company may have a large number of assets such that each scan covers only a subset. In this case, not observing a finding means nothing if the finding occurred on an asset not covered in the current scan. Furthermore, different modes of the same scanner detect different sets of vulnerabilities.

In summary, we uniquely identify a finding by its specific vulnerability, URL or IP, port, and the company associated with the asset at a given point in time. A finding has been resolved when at a later time the same asset range has been scanned with the same tool running in the same mode and the finding is not detected.

The time to fix of a finding, *f*, is calculated as follows:

A given company has been scanned *n* times, with a given scan denoted by s_i. Let s_{first_seen} denote the first observation of a given finding. The finding may then be detected on one or more subsequent scans of the same asset range. Let $s_{missing}$ be the first scan with the same tool on the same asset range that does not observe *f*. The time to fix is thus $s_{missing} - s_{first\_seen}$.

Note that a finding may still be detected in the last scan, s_n, for a company. In this case, we denote the finding's time to fix as '$s_n - s_{first\_seen} +$', to connote that the finding was open for at least that long.

In the time to fix figures in the 'Evaluation' section we include findings that were resolved as well as those that were not resolved after the last scan. We represent the latter as having a time to fix of *Infinity* so that the figures convey both the rate at which findings are resolved as well as what percentage of the findings end up being addressed. For example, if a line in a time to fix figure only reaches $y = 0.8$, then 20% of the findings represented by that line were not resolved during the time period of our dataset.

Finally, we treat a finding being discovered and resolved *n* times over many scans as *n* unique time to fixes.

Caveats

There are several aspects of the dataset that impact our ability to draw conclusions about the time to fix of findings. The first is that the resolution of our analysis is inherently limited by the frequency of scans for a particular company. In the best case, a finding could be observed at s_1 and then fixed right before a scan at s_2, so the time to fix is almost exactly $s_2 - s_1$. However, in the worst case, a finding could be resolved one day after s_2, leading us to calculate the time to fix as $s_3 - s_1$, where s_3 is a subsequent scan that could potentially be months later. Thus, the frequency of scans for a company is an upper bound on the accuracy of our time to fix calculations. Another result of the fact that the time to fix calculations depend on the scan time is that the time to fix ECDF figures presented in the 'Evaluation' section occasionally have large jumps, which correspond to scans.

Finally, there are inherent difficulties in analysing the relative time to fix between companies or industries. Each company and industry has not only a different number of scans but the time between scans may vary, making direct comparisons challenging.

EVALUATION

In this section we discuss insights gained from analysing the dataset described above, including the categories of findings detected in real companies today, the relative breakdown by industry, and how long it takes different industries to address findings. Note that findings marked as *false positive* are not included in any of these figures.

Due to space concerns, most of these figures have been reduced to half page width. The original full size figures may be viewed at [6].

Findings by category

For clarity of viewing, in the following figures we group the thousands of specific types of findings into the categories described in the 'Methodology' section.

Across all companies

In Figure 1, we include all (true positive) findings in our dataset, including when the same issue is discovered on subsequent scans.

One can see that across all industries, companies most often face issues related to the server or language running their web applications as well as PKI/SSH/SSL, with web application best practices being a distant third. Cross-site scripting (XSS), while by no means the most prevalent issue, was observed almost 15,000 times in our dataset.

Industry findings by category

Note that because there is a wide range in the number of findings for different industries, *the y ranges for the charts are different* so that the relative prevalence of issues within an industry may easily be compared. Normalizing the figures across industries is challenging as industries were scanned a different number of times and in some cases are represented by a different number of companies.

Here we include figures for just four of the industries, the rest may be reviewed in Appendix B. We first discuss the financial services and IT industries (Figure 2), two industries we would expect to have both the motivation and institutional support to have strong security programs.

Interestingly, the top four categories of issues for both companies are: general network issues, PKI/SSH/SSL, web application best practices, and the web application server or language used. As mentioned above, note that the *y* scale for each figure is different.

Companies in the leisure & media industry (Figure 3) have the most evenly distributed types of findings of any industry.

In contrast, the transport industry (Figure 3) has by far the greatest category skew – nearly all of its findings relate to the web application server or language.

We note that the PKI/SSH/SSL category is in the top four most prevalent categories for every industry.

Time to fix

We examine the time to fix of various findings, both across all companies as well as by industry, analysing if the time to fix is affected by the type of vulnerability, its severity, or its solution.

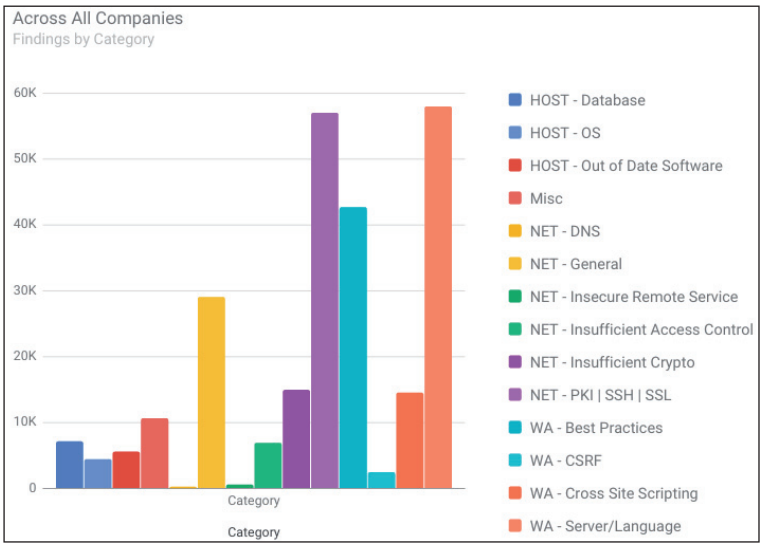


Figure 1: All (true positive) findings in our dataset, including when the same issue is discovered on subsequent scans.

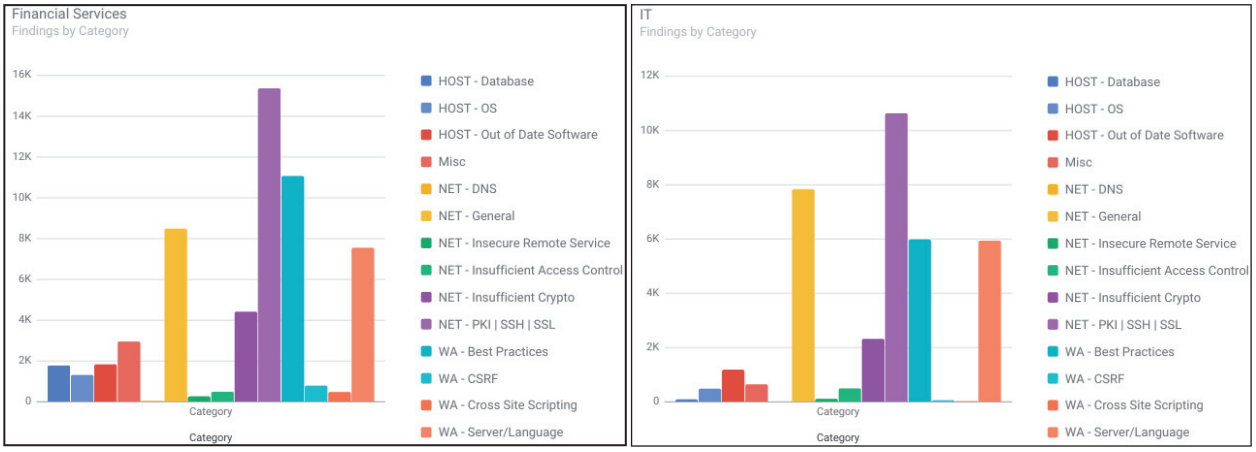


Figure 2: Results for the financial services and IT industries.

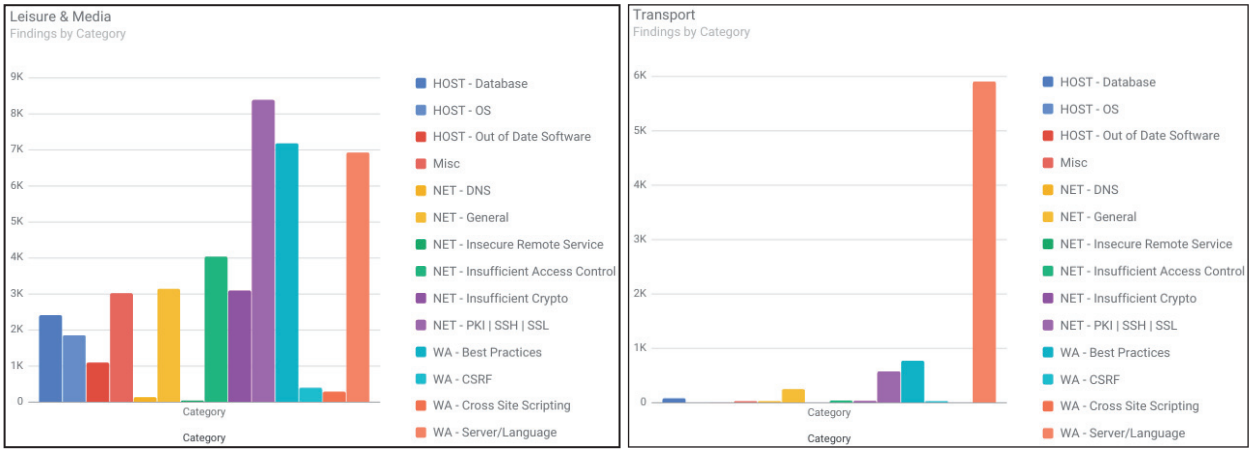


Figure 3: Results for the leisure & media and transport industries.

We use an empirical distribution function (ECDF) to analyse the time to fix data. Each line is 'a step function that jumps up by *1/n* at each of the *n* data points' [7]. In other words, a point *(x,y)* on a line represents that *y%* of the data the line is based on is less than or equal to *x*. For example, the point *(20 weeks, 0.5)* on a time to fix line represents that 50% of the findings in this category were resolved within 20 weeks.

In many of the ECDF figures the lines do not reach $y = 1$, indicating that a percentage of the findings were not resolved within the timeframe of our dataset.

By CVSS

We first examine the time to fix of findings by severity, as denoted by CVSS score. We find that as expected, over time, a greater percentage of higher CVSS score findings (5+) are resolved. However, there is not a complete correlation because findings with a CVSS score in the range 3–4 are actually less likely to be resolved than those with CVSS score of 1–2, by about 10%.

CVSS score does *not* appear to have a direct impact on the speed with which a finding is resolved, as in the first 10 weeks, findings with a score of between 3 and 6 are more likely to be resolved than those in the range of 7–10 (Figure 4). One potential explanation for this behaviour is that findings with a lower CVSS score are easier to address and are thus resolved quickly, while more serious findings require more effort and are thus addressed more gradually.

We next examine the finding time to fix of specific industries by CVSS score group, first of findings with a CVSS score of 7 or greater (Figure 5) and then 5 or greater (Figure 6). To prevent the figures from being overly crowded, we split the industries into two groups and create an ECDF for each.

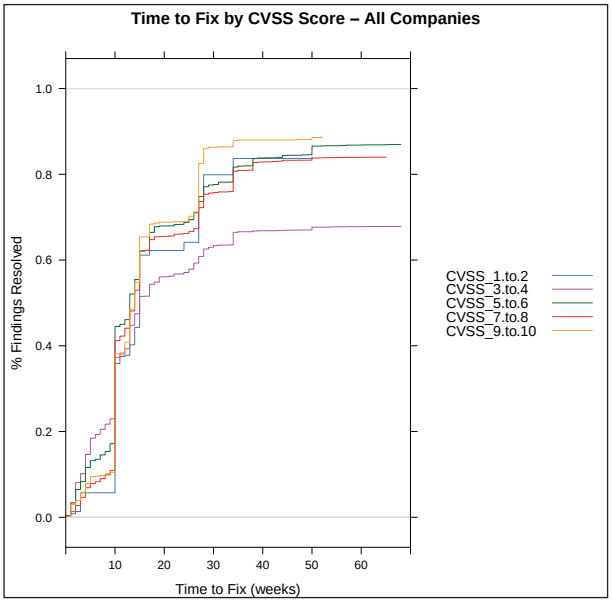


Figure 4: Time to fix by CVSS score (all companies).

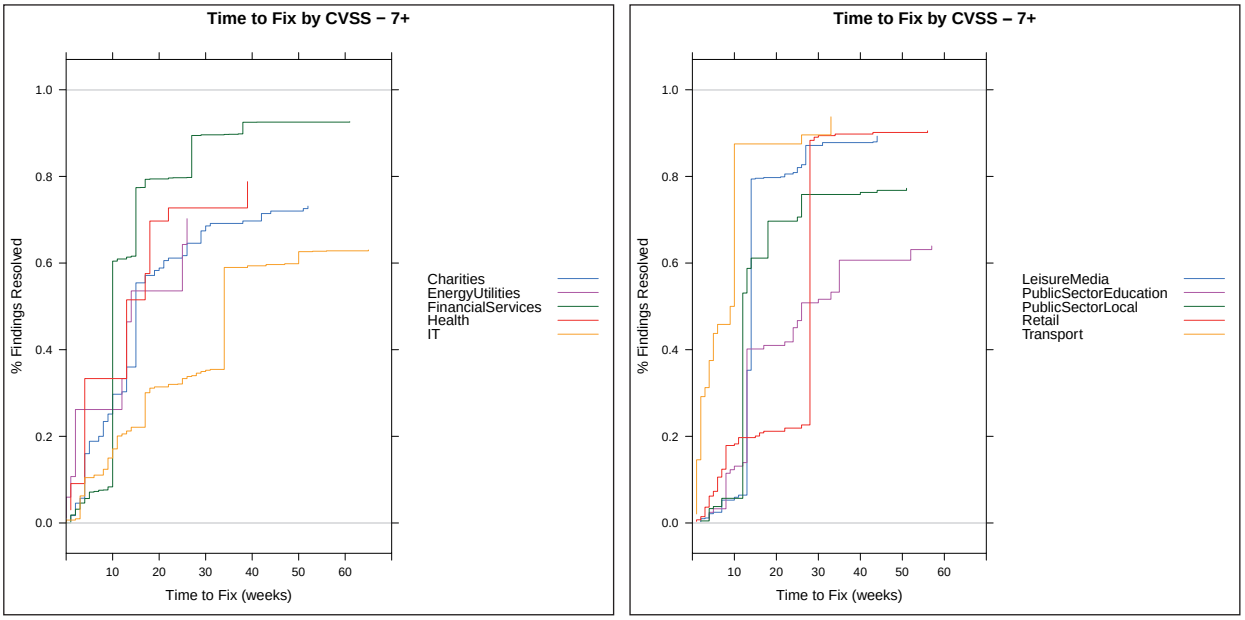
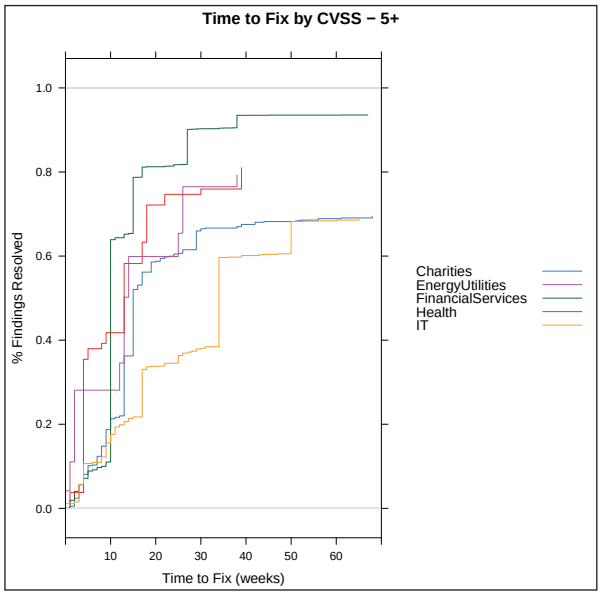


Figure 5: Time to fix with a CVSS score of seven or greater.



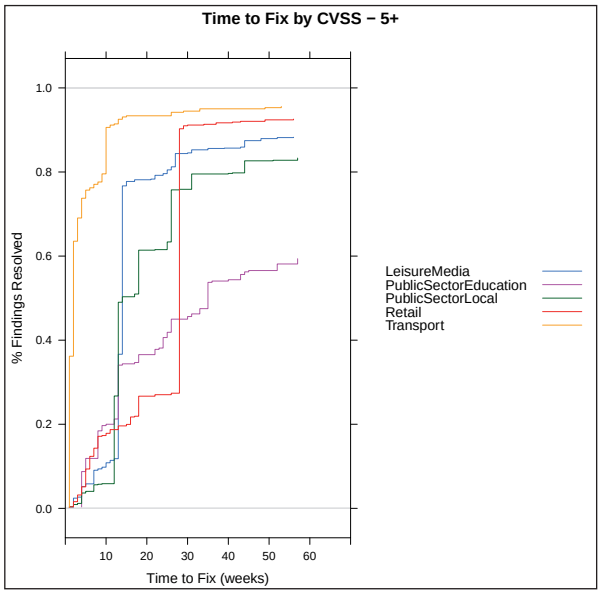
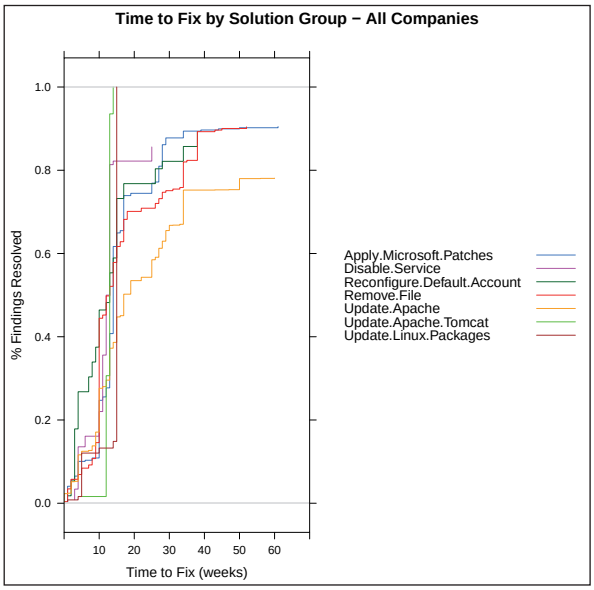
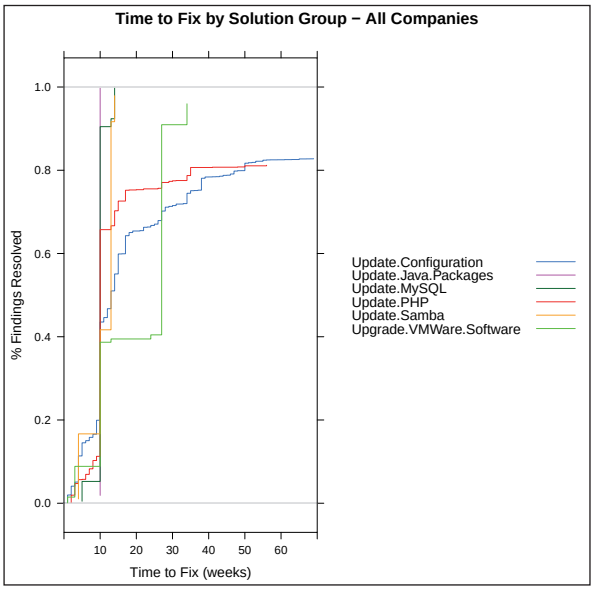


Figure 6: Time to fix with a CVSS score of five or greater.





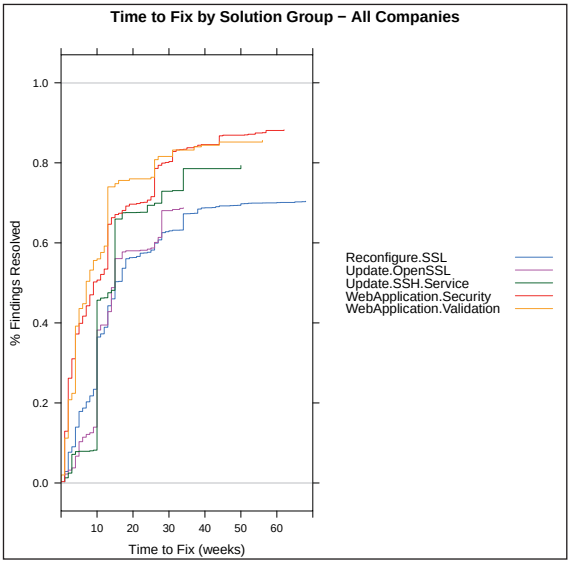


Figure 7: Time to fix by solution group.

If companies prioritized findings with greater CVSS scores, the lines in the figures representing CVSS scores of 7+ would be steeper overall than those in the 5+ figures – a greater percentage of findings would be resolved in less time. However, one can see this is generally not the case and in fact the transport industry in particular resolves findings with a CVSS of 5–6 much faster than just the findings in the 7–10 range.

By solution group

We examine whether the time to fix of findings is influenced by the remediation required, as denoted by the *solution groups* associated with a finding (Figure 7), as described in the 'Dataset' section.

- We find a significant difference in the time to resolve OS findings – 100% of *Linux* packages were updated within 20 weeks, while only about 75% of *Microsoft* patches were applied in a similar time frame, and 10% of *Microsoft* patch issues were still unresolved after a year.
- In general, a high percentage of findings that require updating a language or package tend to be resolved, usually within 20 weeks, except for *Apache* and PHP.
- Findings that require more environment-specific solutions, such as disabling a vulnerable service, reconfiguring default accounts, removing files and updating configurations tend to take longer to be resolved and are more likely to be left unresolved.
- Cryptography-related issues, such as those requiring reconfiguring SSL or updating OpenSSL or SSH have the lowest likelihood of being resolved, ranging from about 70%–80%.
- Over half of web application-related findings were resolved within 10 weeks, though more than 10% of them were never resolved.
- Across all solution groups, we observe that the majority of the findings that end up being resolved are addressed

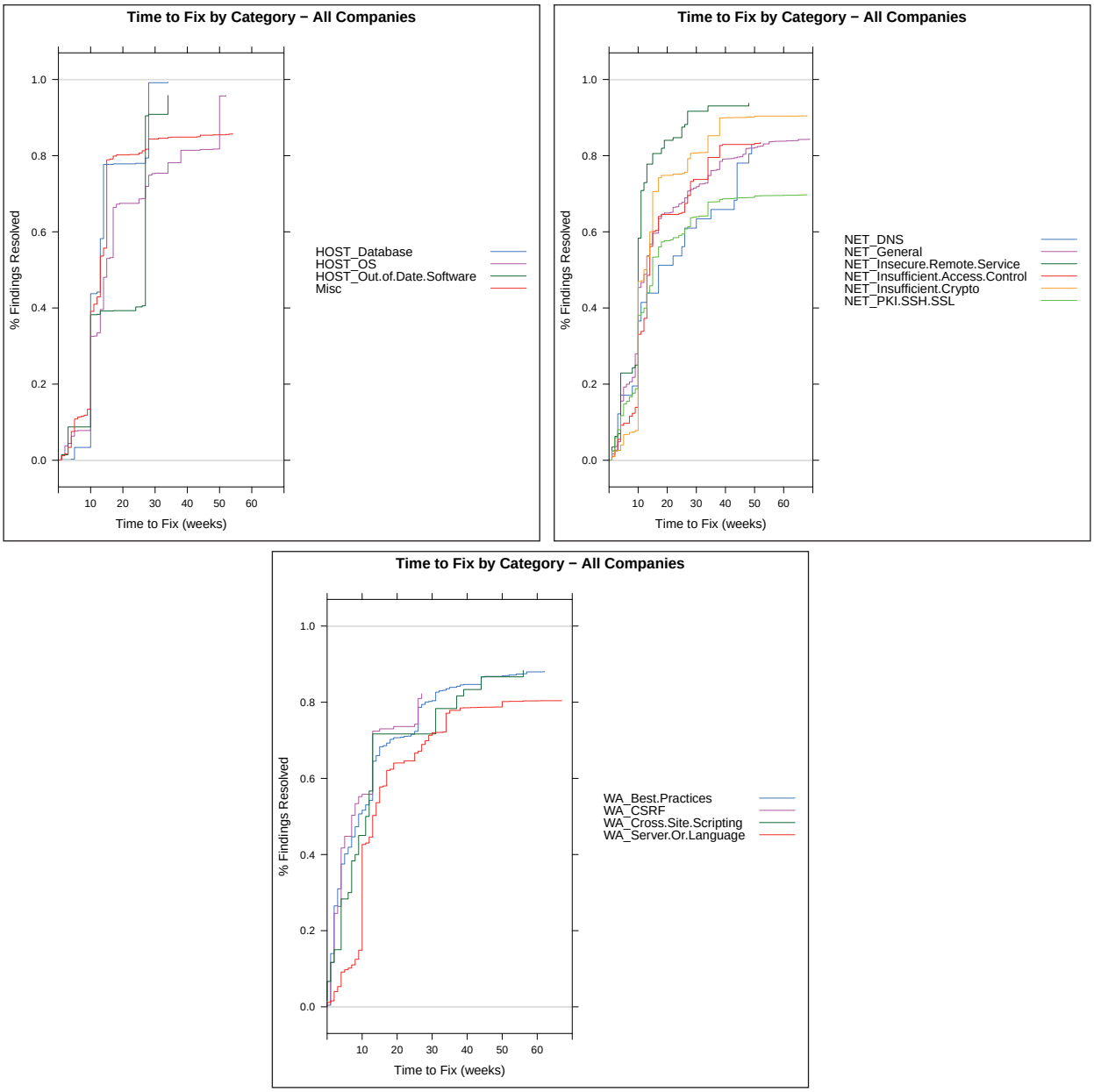


Figure 8: Time to fix of findings across all companies, grouped by category.

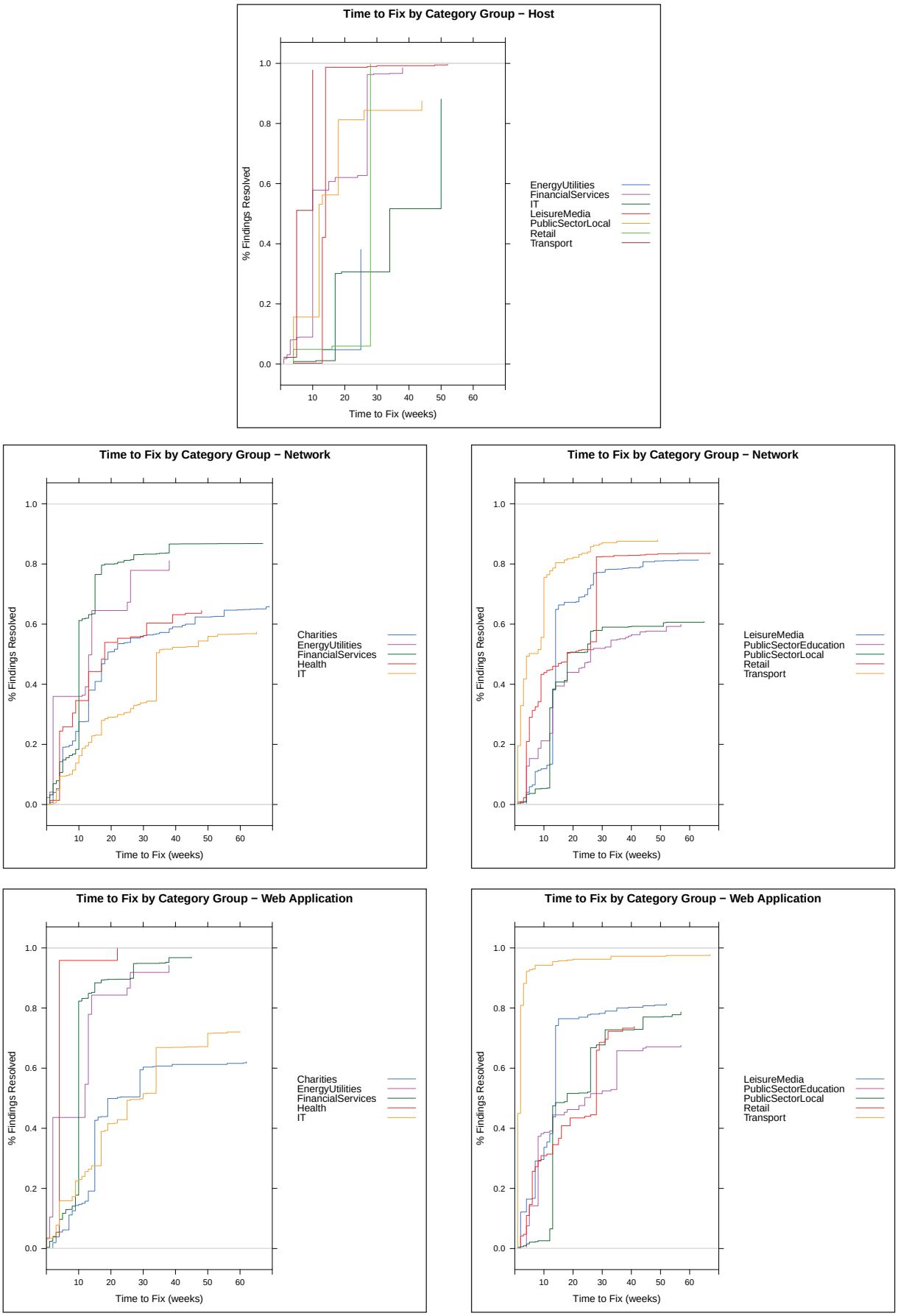


Figure 9: Time to fix by category group.

within the first 10–20 weeks, generally over 50% of the findings, with the rest resolved more slowly.

- We feel this intuitively makes sense, as an organization will likely make an initial large push towards addressing the security issues they are informed about. However, the issues that are not resolved in the initial push are addressed slowly, when time permits, over the subsequent months.
- The complexity of the solution, as mentioned above, is likely also a factor.

By category, all companies

We examine the time to fix of findings across all companies, grouped by category (Figure 8).

We see that host-based findings, which tend be addressed by package updates or applying patches, are often tackled in batches, leading to the sharp cliffs. Host-based findings are resolved at the highest rates, all over 90%, though out-of-date software findings are resolved the slowest of any category, with only about 40% of issues being resolved after 30 weeks.

Companies appear to make a big push to resolve network-based findings in the 10–20 week time frame. Interestingly, the DNS category is one of the few categories in which a large percentage of the findings that end up being resolved are addressed *after* the first 20 weeks. We believe this may be due to increased care being needed to address DNS-related issues in a way that maintains service uptime. PKI/SSH/SSL issues are resolved overall at the lowest rate, about 70%.

Different categories of web application findings tend to be resolved at a similar rate. The majority of the findings that end up being resolved are addressed within the first 20 weeks; in fact, no CSRF findings are fixed after 30 weeks though all of the other types continue to slowly be addressed past 50 weeks.

By category group and industry

We next examine each industry's rate at addressing categories of issues, grouping all host, network, and web application findings together (Figure 9). Note that we filter out industries without a threshold number of findings in that category, so not all industries are represented in every set of figures.

We see a large disparity between industries in both the time to fix and overall percentage resolved of host-related findings. The transport industry resolved nearly all of its host findings within 10 weeks and the leisure & media industry soon after, while it took the IT industry around 35 weeks to resolve just 50% of its findings and the energy & utilities industry resolved less than 40% of its host-related findings in total.

No industry resolved more than 90% of their network-related findings, and half of the industries resolved around 60% of their findings. Overall, the transport industry resolved its network findings the most rapidly, and along with the financial services industry, had the highest percentage resolved.

There is also a sizable disparity both in the time to fix and overall percentage resolved of web application findings. The transport and health industries resolve over 90% of their findings in under 10 weeks, while the charities, IT, public sector and retail industries are still resolving many of their findings in the 30–50 week time frame. One surprising trend we note is that IT is one of the slower industries at resolving web application findings and that it only resolves about 70% of them.

Of all three groups, host-based findings are resolved at the highest rate across industries, followed by web application and lastly network-related findings. As one might expect, the financial services industry resolved one of the highest percentages of findings in each of the categories, though not necessarily the most quickly.

By specific category

We examine further how each industry handles PKI/SSH/SSL (Figure 10) because it is one of the most prevalent finding categories across all industries and because of its importance.

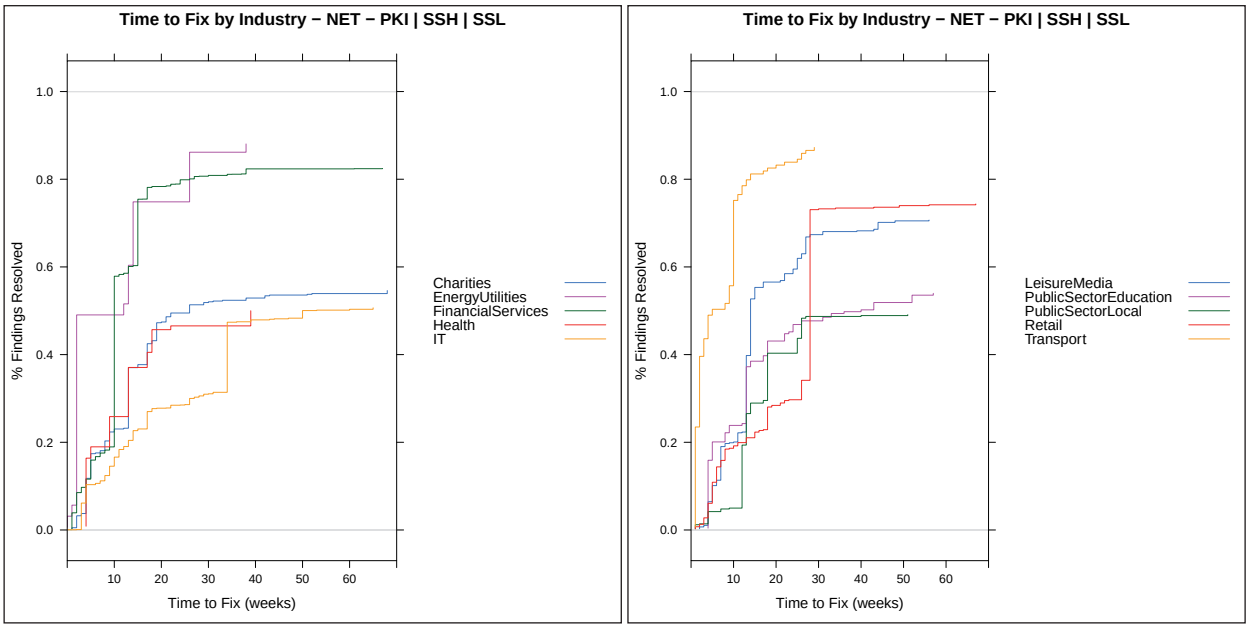


Figure 10: Time to fix PKI/SSH/SSL.

It appears that companies have difficulty resolving findings in this category – no industry resolved more than 90% of their issues and half of the industries only resolved about 50% or less. Again, the transport industry performed well both in fixing their issues rapidly as well as addressing a high percentage of them, with the energy & utilities and financial services industries also performing well.

CONCLUSION

Technology continues to advance at a rapid rate, and with increased focus from determined attackers, organizations are feeling more pressure than ever to keep their websites and infrastructures secure. Due to time and resource limitations, many organizations are leveraging automated security scanning tools to find misconfigurations, services that need to be patched, and web application vulnerabilities.

In this paper, we present insights gained from analysing the results of running a security scanning tool on 100 international companies across 10 industry verticals from February 2014 to May 2015. We examine the types of findings discovered both by industry and across all companies. We discuss the time to fix of findings by category (broken down into subcategories of host, network and web application), severity (as denoted by CVSS v2 score), and remediation solution.

We hope that this work will provide industry professionals with objective data against which they can compare their companies' performance, and security researchers with insights into impactful areas on which they can focus in their future work.

REFERENCES

[1] Golden, H. (2015, April 16). Shortage of IT Security Professionals Not Unique To Government. http://www.nextgov.com/cio-briefing/wired-workplace/2015/04/calling-all-information-security-professionals-world-needs-you/110338/.

[2] SANS. NetWars. https://www.sans.org/netwars/.

[3] USCC. US Cyber Challenge. http://www.uscyberchallenge.org/.

[4] Risk Based Security. The CVSSv2 Shortcomings, Faults, and Failures Formulation. https://www.riskbasedsecurity.com/reports/CVSS-ShortcomingsFaultsandFailures.pdf.

[5] Neohapsis Labs. CVSS – Vulnerability Scoring Gone Wrong. http://labs.neohapsis.com/2012/04/25/cvss-vulnerability-scoring-gone-wrong/.

[6] Gibler, C. VB2015. Examining Corporate Security Practices Supporting Figures. https://github.com/clintgibler/vb2015_examining_corporate_security_practices.

[7] Wikipedia. Empirical Distribution Function. https://en.wikipedia.org/wiki/Empirical_distribution_function.

APPENDIX A – CATEGORIES

In this section we give several illustrative examples of the types of vulnerabilities that are mapped to the high-level categories used in this paper. As there are over 9,000 unique vulnerabilities, an exhaustive list is space prohibitive.

- Network
  - *General*: miscellaneous network issues, for example findings related to *Samba*, WebDAV, SMTP, UDP, HTTP TRACE, etc.
  - *Insecure remote service*: unencrypted Telnet available, VNC's remote control service installed.
  - *Insufficient access control*: CIFS share world readable or writeable, default Telnet or SSH passwords, FTP access with standard credentials.
  - *Insufficient crypto*: credentials for a service being transmitted in cleartext, SMB signing disabled or not required, cryptography-related implementation weaknesses in the configuration of a network service.
  - *PKI* | *SSH* | *SSL*: vulnerable SSH or SSL versions being used, weak ciphers or keys, Heartbleed, X.509 certificate is invalid/expired or the certificate subject CN does not match the entity name.
- Web application
  - *Best practices:* e.g. not using the secure flag on cookies served over SSL, not using the X-Frame-Options header (a clickjacking mitigation), form actions submitting sensitive data in the clear.
  - C*ross-site request forgery (CSRF)*: all types.
  - *Cross-site scripting (XSS)*: all types – standard, DOM-based, etc.
  - *Server/language*: related to the server-side language or server itself, such as PHP issues or *Apache/Microsoft IIS* CVEs.
- Host
  - *Database*: database-related findings – *Oracle*, *MySQL*, etc.
  - *OS*: issues/CVEs specific to the host OS, including *Solaris*, *Linux*, *Windows*, etc.
  - *Out of date software*: obsolete version of the host OS or host software, including *Microsoft Office*, sendmail, and BIND.
- Misc
  - A wide variety of issues ranging from general CVEs to guest access being allowed for *Windows* logs, and anything that does not naturally fit into one of the above categories.

APPENDIX B – INDUSTRY FINDINGS BY CATEGORY

The following are the industry findings by category figures that were omitted from the 'Evaluation' section: