

META-CYBER SECURITY SYSTEM

Our increasingly complex digital world requires a combination of techniques for cyber security monitoring, including effective graphical representations, providing the holistic solution to the network security Big Data problem.

(The original version of this article appeared in issue 22 of Digital Forensics Magazine for February 2015).

Daniel S. Fowler, Madeline Cheah, Bob Bird
Coventry University, Priory Street, CV1 5FB

ABSTRACT

Digital security and forensic specialists tread water in a sea of data searching for the digital droplets that pinpoint suspicious activity. Technology is the cause of the increased complexity in our connected digital systems, and technology must be used to handle this complexity. Several factors detrimental to cyber security systems efficiency have been identified. A sub-set of the factors affects the responsiveness of the human operators. Low performance and low efficiencies in cyber security systems increases breach risk and loss of service risk in the digital network infrastructures and attached devices. To restore performance and efficiency a software framework is proposed that uses a combination of techniques to address scalability issues combined with clear information visualisation. This framework, whilst initially used as a meta-cyber security system, is adaptable to other big data analysis problems.

INTRODUCTION

It has taken one generation for the world to become enmeshed in a digital web. In the global multi-trillion dollar digital playground organisations run their businesses, people organise their lives, entertainment is on tap and new media replaces old. With smartphones in our pockets and tiny network enabled computers embedded into everyday items; computing can already be considered pervasive. Yet wherever society treads, crime follows, and for every technological invention created to aid humankind, criminal uses of the same technology soon evolve. Internet-based crimes, and the cyber security systems to combat them, are almost as old as the Internet itself. Existing criminal activities are now performed in new and ever more creative ways, and criminals operate across borders at low risk for high reward.

The increasing difficulty in protecting large scale digital networks

Despite the existence of well-established cyber-security systems such as intrusion detection and prevention systems (IDS/IPS), digital networks and the devices connected to them are compromised with alarming regularity. The supposition of if a network is breached has been replaced by the presumption of when it is breached. It begs the question why? The phenomenal success of digital technologies in this "Information Age" can also be considered an Achilles' heel. Continuous innovation sees digital devices and networks scaling ever upward. Data transmission rates multiply and computer processors exponentially increase their processing power. There is more data storage space on quicker hard drives supporting the huge volumes of data travelling on the networks. The incremental progression towards cloud computing continues as more applications and services migrate online. The number of connection points continues to increase with the International Telecommunication Union (ITU) estimating that 60% on the World's population will be online in 2015. These advances help society but also enable and empower criminal activities. They use the technology, building bigger and faster botnets whilst using the anonymity of the Internet to hide and collude. They hunt for the bugs, backdoors and vulnerabilities in the complex software systems that run the digital world. They trap the untrained public into revealing personal and security details in phishing attacks. If the insider threat from espionage or malicious employees is not a concern for an organisation it should be.

Why do cyber security systems fail to identify and stop breaches? Research has identified recurring themes:

VOLUME

Network attacks are relentless; once a device is connected to the Internet, it does not take long before they are probed, usually by automated systems. Because of this, intrusion detection and prevention systems (IDPS) tend to generate a large number of alerts and false positives. This is exacerbated by the complexity of configuration, where incorrect configurations, imprecise rules and even normal day-to-day traffic add to the sheer volume of data to be sifted through, sometimes manually.

SCALABILITY

Network security scalability concerns have been around for some time (Shaikh, et al. 2009). Despite this there has been little research into scaling cyber security systems for large networks and data volumes, and doing this in an intelligent and efficient manner. Such research would provide new perspectives on the issues highlighted here.

TESTING

Quantitative testing of cyber security systems is difficult. There does not exist any standardised modern large scale datasets or test methodologies. Old datasets, used by some researchers, can be acquired but will not replicate modern attack methods as data patterns and the infrastructure used to carry this data changes quickly. The reliability of these datasets, whether for training or simulation can also be questionable.

RESOURCES

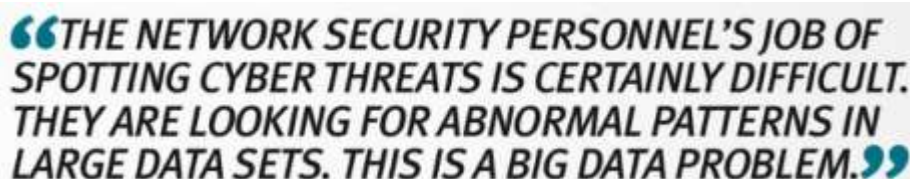
There is a limited pool of skilled personnel capable of performing cyber security and network forensics roles. Personnel may be performing such roles alongside other duties and thus have time restrictions when dealing with issues.

Finally, cost, as always, is a major consideration. Security is a necessary overhead in any organisation; systems that are time consuming to deploy, maintain and administrate and are resource heavy are not viewed favourably. On the flipside any system that mitigates costs would be welcomed.

If data and alert volumes, and the lack of resources to deal with them, are causing organisations headaches how can the issues be addressed? The use of digital networks and systems is accelerating with the advent of the Internet of Things. Thus it becomes a case of fighting fire with fire: technology is the source of the problem and it must be applied to find the solution.

VISUALISATION

The network security personnel's job of spotting cyber threats is certainly difficult. They are essentially looking for abnormal patterns in very large data sets - in other words - a big data problem. The interconnected society is producing massive datasets that are being analysed for trends, statistics and information, often with the view to gaining a competitive edge. As a result visual analytics, defined as informational visualisation combined with analytics, is big business, with a wealth of new tools and techniques available which can be harvested for ideas to deal with big network data visualisation.



“THE NETWORK SECURITY PERSONNEL'S JOB OF SPOTTING CYBER THREATS IS CERTAINLY DIFFICULT. THEY ARE LOOKING FOR ABNORMAL PATTERNS IN LARGE DATA SETS. THIS IS A BIG DATA PROBLEM.”

Cyber security data visualisation is a well-established and active field. There are a plethora of visualisation techniques applied to network data for management and monitoring and many current IDPSs have visual displays. Development work is always underway however, and the latest research papers show novel visualisations of network and security data (Shiravi, 2010). Despite all this it is difficult to find a successful, widely used, go-to cyber security visualisation product. Many visualisations are not generated in real-time and display static or temporal summary charts from network packet captures. There are others that are overly

complex or badly designed and there are packages with representations that are difficult to interpret and thus defeat the whole idea of using visualisation. What is needed is the application of current visual analytical practices applied to cyber security's big data problem to improve real-time situational awareness.

What is "Cyber-Situational Awareness"?

Many advancements in technology are as a result of state spending on research for military purposes. Research into fighter pilot and astronaut cockpits led to investigations into technology that could improve "situational awareness". With the aim being to allow for pilots to rapidly process the information related to several complex environments: the aircrafts controls and flight parameters, the weather, the position of hostile and friendly aircraft, mission objectives, terrain, ground units and radio communications. Situational awareness began to appear in research papers in the 1980s dealing with visual displays that allowed for rapid cognitive processing and information assimilation, supported by NASA and the US air force. The phrase was taken up by other military, medical and law enforcement organisations, wherever complex environments needed an organised approach to handling multiple sources of information. The complexity of monitoring an organisation's large digital network and the myriad of devices connected to it warrants the use of software to provide situational awareness of systems status, network status, status of services, log file management, vulnerability and patch management, backups and cyber security events.

The essence of great visualisation is to keep things as simple as possible. Poorly designed gauges and cluttered graphics take up valuable space, distract the user and obfuscate the information the data is trying to convey. Instead, reliance upon simple graphical elements should be a priority, with attributes applied for "preattentive processing"; defined in the science of psychology as the process of decoding visual information without conscious thought. The lack of conscious thought results in high cognitive speed. There is also no need to reproduce an accurate physical representation of real world devices on the screen. Personnel know what computers, routers and switches look like, simple abstract representations are more than adequate and less distracting.

38753542510442717843
06342838894799866014
91026705452294623532
03452985830191630250
77322917150051834565

387**5**3**5**42**5**10442717843
06342838894799866014
9102670**5**4**5**2294623**5**32
034**5**298**5**8301916302**5**0
773229171**5**00**5**1834**5**65

Figure 1. How many fives? Preattentive size and intensity attributes on the bottom lines aids cognition.

Displaying Data – A Manual

The book "Information Dashboard Design" by Stephen Few (ISBN 978-1938377006, Analytics Press, July 2013) is appropriately subtitled "Displaying Data for At-a-Glance Monitoring". It is a go-to manual for anybody involved in designing a visual analytic display for complex data. Big data analytics is becoming increasingly important in cyber security systems and digital forensics and this book shows how to design effective visualisations, especially with regards to using preattentive processing effectively.

Quantitative Testing Of Visualisations

The quantitative testing of cyber security visualisations is thin on the ground and would certainly benefit from additional research. In 2005 a paper by Oracle showed the advantages of a treemap visualisation compared to a traditional tabular format for monitoring network data. In a recent paper from Edinburgh Napier University on the design of an IDS visualisation tool a test was performed on its effectiveness compared to traditional command line tools. Using the NASA Task Load Index criteria to evaluate the analyst's workload compared to command line tools the visualisation tool was more effective (the lower the average score for the dimension the better). Whilst this was only a small study group such studies for other type of visualisations would be beneficial to the cyber security and digital forensics community.

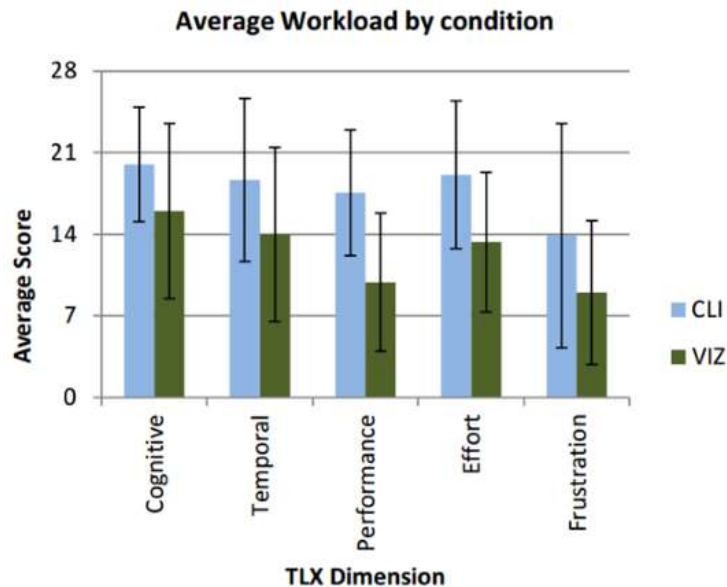


Chart 1. Thomson, A.; Graham, M.; Kennedy, J.: "Pianola - Visualization of Multivariate Time-Series Security Event Data," Information Visualisation (IV), 2013 17th International Conference, pp.123-131, 16-18 July 2013

Even the Big Guys Fail

Breaches are a daily occurrence in every type of industry and organisation - <https://www.hackmageddon.com/>

- Nov. 2014 - Another Sony hack reveals high volumes of sensitive corporate data.
- Aug. 2014 - JP Morgan Chase, data on 76 million households and 7 million businesses is compromised.
- May 2014 - eBay breach, 145 million users recommended to change passwords.
- Oct. 2013 - Adobe Systems: 152 million customer records stolen.
- April 2011 - Sony breached, data from 77 million user accounts revealed.
- Jan 2007 - TJX, 94 million credit card numbers stolen.

AN HOLISTIC APPROACH

Taking into consideration the above, the scalability problem can then be addressed by taking an holistic view of the large complex digital networks that organisations now use, a "visual of the visuals" as it were. Enter the meta-cyber security system: a cyber security system for cyber security systems. This meta-system does not replace existing well-established IDPSs, it uses them as a component in this holistic system and uses a divide and conquer approach to address scalability limitations. The outer IDPSs on a large network processing the traffic on their segment and alerting the central cyber security system.

A consensus view that emerges from the research is the need for an automated or semi-automated cyber security system to handle the high volume of alerts. Taking suggestions from previous research and combining those techniques with visual analytics it is possible to design a system to provide real-time situational awareness, automated alert filtering and optional automated or semi-automated cyber security tuning. This would ease the workload of network analysts and improve their efficiency.

This sounds not dissimilar to a Distributed Intrusion Detection System (DIDS). Such a system aggregates the alerts from outlying security systems. The alerts are collated and filtered according to a data fusion algorithm. A visual display is used to convey system status and allows personnel to drill down into the data. This aggregation, collation and display of multiple alerts is similar in functionality to the aggregation and collation of log files in log and event management systems. Some DIDS designs support automatic tuning, for example updating a firewall rule to block traffic from a specific device if malicious activity from that device is spotted.

Yet this design has flaws, as some of the research reveals. Collecting, collating, processing and displaying alerts from outlying security systems could magnify issues that are present in all cyber systems, namely the limitations on processing power (CPU throughput) and data storage space; a busy 1 Gbps office network can generate several terabytes of data a day. At that rate any storage devoted to capturing network data for forensic analysis is soon depleted.

So where does all this information come from?

The process of identifying patterns within data sets can be achieved using data mining. As more volume is accumulated and stored, data summarisation (for example, through visualisation) can be used to establish a baseline and identify anomalies, patterns, trends and key events. Just as importantly, organisations can then act upon this data in an informed manner. A word of warning: learning cannot take place without assumptions! Naturally, it follows that the premise of such assumptions would have to be as sound as possible, otherwise everything else built upon it may fall.

MITIGATING CPU THROUGHPUT

In an IDS or IPS processing time is used to analyse the network traffic. If each unit of work takes too long then network packets are dropped and the security system is not providing 100% coverage. This applies to the aggregation and collation of alerts from several systems. It needs to be quick enough to provide 100% coverage. A central system processing alerts from outlying systems is not handling real-time network traffic. However, it must still be quick enough to handle the alerts from all the outlying security systems, especially during an alert storm. This is when a cyber attack causes a large number of alerts to be generated in short timescale. To improve processing throughput proposals include:

1. Only perform simple analysis, but the simplicity may increase false positive rates, whilst complex or multi-stage attacks may escape undetected.
2. Using parallelism in code by using multi-threaded code running on multi-core processors allows for execution of more units of work in any given timescale.
3. Using graphical processing units (GPU) cards that have hundreds of CPU cores. These employ processing techniques similar to the divide and conquer techniques used for image processing and is another way of achieving parallel processing.
4. Building cyber security features into the core functionality of networked devices to minimise processing delays. For example, a version of Berkeley Software Distribution (BSD) has a packet filter integrated deep into its network stack to reduce propagation delays when filtering network traffic.
5. The installation of statistical and behavioural analysis techniques in addition to detailed packet analysis to examine the overall traffic patterns. However, this could increase false negatives.
6. Compiling IDS and IPS rule sets into machine code. An example of this is to encode rules into C code, compiling and then linking the code directly into an executable, or running it on optimised hardware such as GPUs or a Field Programmable Gate Array (FPGA).

THE STORAGE PROBLEM

Ideally all data processed by a security system is stored for future possible use in forensic investigations, management or compliance reports. Daily network captures on large busy networks requires terabytes of storage capacity but economic and physical constraints dictates that storing every piece of network data is not

practicable. Furthermore, protections for this data in order to assure integrity would further add to the computational overhead.

Cyber security systems need to employ sophisticated data filtering techniques and only store relevant information. The definition of what is relevant or not is a problem familiar to personnel handling system log files. Best practice handling techniques can be applied, including file rotation, short retention periods and archiving to other external media. The problem with data filtering techniques is that it may impact upon information required to investigate the source of the attacks, the need for evidential integrity and the storage of data for compliance with regulatory bodies and statutes.

A central cyber security system cannot be responsible for storing all the data from the outlying systems, because of the data volume issues. It needs to store the outcome of its own processing, relying upon the security systems that report to it to store the detailed data or off-load data storage to dedicated disk arrays.

The meta-cyber security system is based upon a flexible component based framework that can be adapted to the different requirement that organisations will have. No two networks are the same so the system needs to be adaptable. A network communications module talks to the outlying cyber systems receiving the IDS and IPS alerts and other event information (log file and process monitoring events) and will support syslog formats over UDP, XML or JSON over HTTP and file transfers. Alerts received from the outlying systems may require normalising for date and time as well as data formats.

At the heart of the design are modules similar in nature to those used in a DIDS for the aggregation, collation and reporting of alerts, performed in real-time by taking advantage of multi-core processors and database technology. Research papers have proposed several algorithms for the aggregation and collation of cyber events. These algorithms include: data clustering, decision trees, data fitting, state transition analysis, principle component analysis and a long list of machine learning methods including neural networks, fuzzy logic, bayesian networks, expert systems, maximum likelihood estimation, generative modelling and descriptive modelling. Rather than choosing a specific algorithm the component based design of the meta-system allows for different algorithms to be added as they are developed. It comes initially with a rules-based engine to allow for the implementation of decision trees. The rules engine not only drives the algorithms for the alert aggregation and collation but also provides the rules for the real-time visualisations and optional tuning. The rules are stored in a database and can be added to and modified using a rules editor.

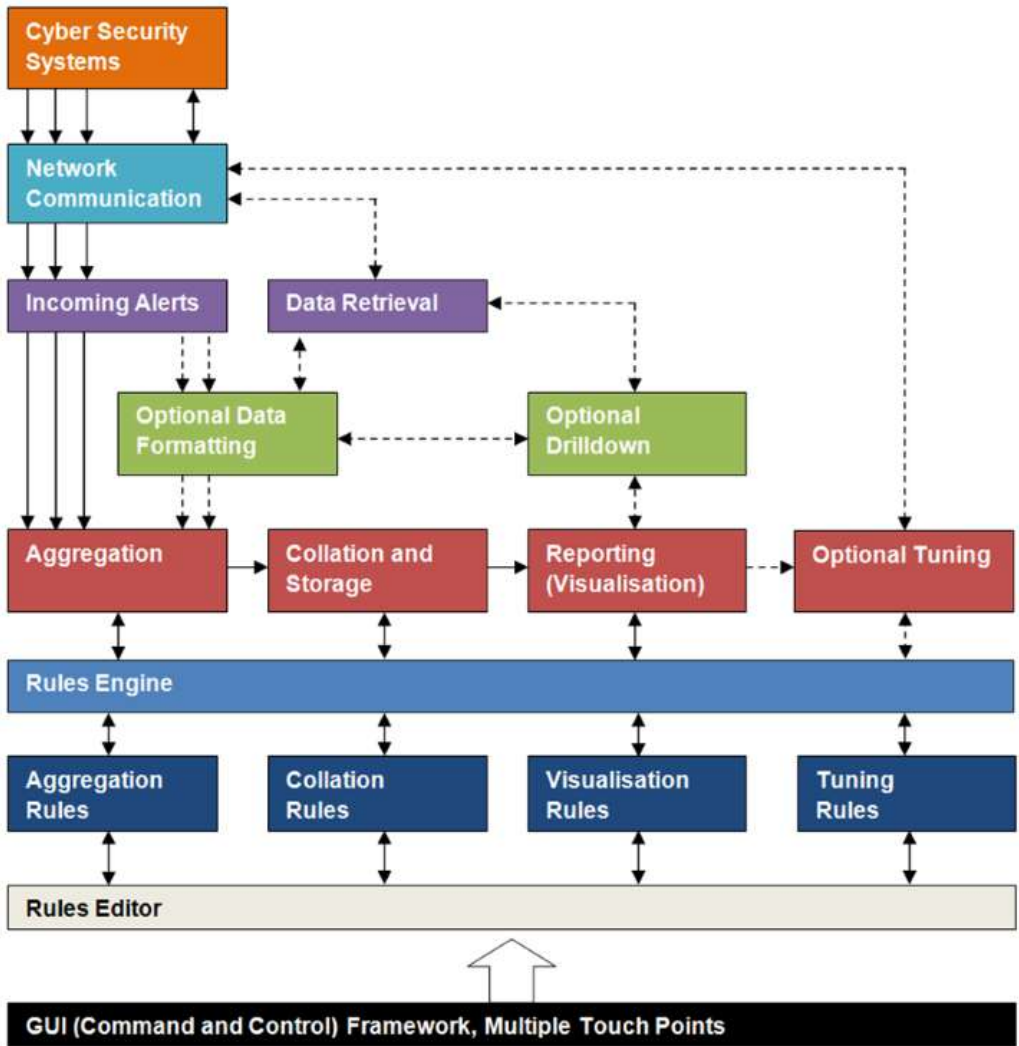


Figure 2. The major logic components of the Meta-Cyber Security System

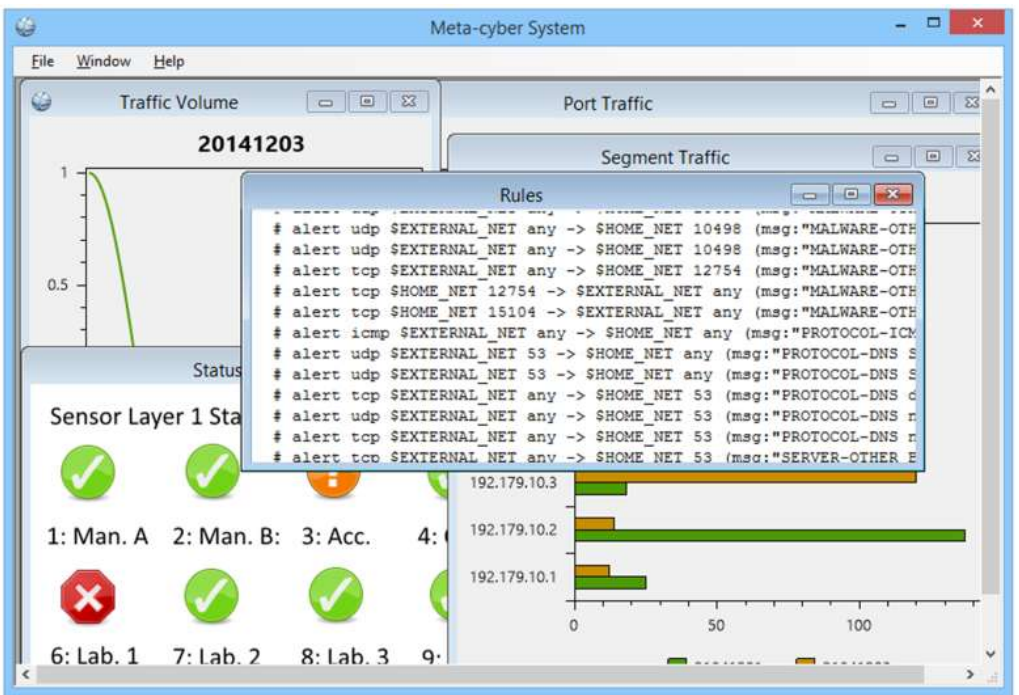


Figure 3. A Meta-Cyber Security System must be adaptable to fulfil an organisation's requirements

A graphical user interface (GUI) provides the command and control interface to the system. The GUI hosts the cyber events and visualisations from the reporting module (summary views and drill down views), the rules editor and system management functionality (shutdown/restart and database maintenance).

The design of this meta-cyber security framework is adaptable to other big data problems. In essence it allows for algorithms to be executed across several data streams to search for specific data patterns and highlight those patterns on an interactive display. Those data streams need not be restricted to network security data. Scalability issues exist in other technology areas: the huge datasets found on multi-terabyte hard drives, in cloud computing systems, open data sources and big business databases.

Network attacks occur in real-time so the system is designed to process alerts as they occur, informing the operators of events as they happen. Running the system offline provides the ability to be able to replay events for forensic analysis and intrusion investigation.

8TB a day

The computer networks in organisations have increased in data transmission rates from 10 Mbps through 100 Mbps to 1 Gbps. On a busy network, 1 Gbps could see approximately 100 MB travelling between devices every second: that is 6 GB per minute! It is possible for the network to handle over 8 TB every 24 hours. At that rate the data storage set aside for packet capture is soon exhausted. Data reduction and management techniques that reduce the volume but preserve evidence of cyber attacks are required.

What is the biggest mistake to avoid for data visualisation?

Assuming that the sum of all knowledge is held within the dataset itself. Datasets can be biased, sometimes intentionally, and meta-data should be considered within this, who acquired the data? Where was it stored and who has access to it? How was it gathered and why? Any evidence that may refute, complement or supplement the data should also be taken into account.

CONCLUSION

The high profile of cyber security attacks and the consequential reputational and economic loss has attracted investment from governments and organisations to improve cyber security products. The national infrastructure reliance that countries have on Internet connected systems, especially in developed countries, has seen increased funding for cyber security initiatives. In the UK the governmental Office of Cyber Security and Information Assurance has some high profile aims and objectives. No system can ever be absolutely protected all of the time, especially when the weakest links are the human users. Although long-standing scalability concerns in cyber security systems exist the technology is available to address them. It is not one technology but a combination of technologies from the fields of Computer Science, Digital Forensic Science, Psychology and Visual Analytics that is the way forward for humans to comprehend the vast volumes of data that travel across the digital networks.

REFERENCES

Siraj A. Shaikh, Howard Chivers, Philip Nobles, John A. Clark, Hao Chen; Towards scalable intrusion detection, Network Security, Volume 2009, Issue 6, June 2009, Pages 12-16, ISSN 1353-4858, [https://dx.doi.org/10.1016/S1353-4858\(09\)70064-9](https://dx.doi.org/10.1016/S1353-4858(09)70064-9)

Shiravi, H.; Shiravi, A.; Ghorbani, A.A., "A Survey of Visualization Systems for Network Security," Visualization and Computer Graphics, IEEE Transactions on, vol.18, no.8, pp.1313-1329, Aug. 2012, <https://dx.doi.org/10.1109/TVCG.2011.144>

AUTHOR BIOS



Daniel S. Fowler is a Chartered Engineer with a successful career in the computing industry. He is currently completing a master's degree in Forensic Computing at Coventry University. Dan's varied career in IT has taken him from programming microprocessors in industrial control equipment to managing teams designing and producing software for the financial services industry.



Madeline Cheah is a PhD research student currently looking into the field of vehicular cyber-security. Madeline started her academic career as an Assistant Lecturer in Digital Forensics and Ethical Hacking and is currently a member of the Digital Security and Forensics (SaFe) Applied Research Group at Coventry University. Current research interests include vehicle forensics, the legal implications of digital forensic processes and cyber security education.



Bob Bird is a senior lecturer on the BSc Ethical Hacking and Network Security degree program, and the MSc Forensic Computing program. Prior to joining Coventry University in 2007, Bob was a Superintendent responsible for Territorial Operations in Coventry for the West Midlands Police. He was involved in a wide number of major criminal investigations into offences of murder, had extensive experience as a firearms and Public Order Commander.