2014

Rapid forensic crime scene analysis using inexpensive sensors

Dan Blackman
Edith Cowan University, Western Australia Police

DOI: 10.4225/75/57b3e8b9fb883

12th Australian Digital Forensics Conference. Held on the 1-3 December, 2014 at Edith Cowan University, Joondalup Campus, Perth, Western Australia.

RAPID FORENSIC CRIME SCENE ANALYSIS USING INEXPENSIVE SENSORS

Dan Blackman
Edith Cowan University, School of Computer and Security Science, Perth, Australia
Western Australia Police, Perth, Australia

Abstract
Network forensics and Network Intrusion Detection Systems (NIDS) have ultimately become so important to corporations that in many cases they have been relied upon to identify the actions of offenders and to provide sufficient details to prosecute them. Unfortunately, as data links on corporate networks have increased to saturation, more information is being missed and even though corporations have spent heavily acquiring loud, power hungry devices to monitor their networks. A more power efficient solution, which consumes less electricity, yet provides the same or better packet inspection is an obvious solution.. This paper discusses a possible solution using a cluster of Raspberry Pis, a credit card sized computer valued at AUD$40 each. These tiny devices (whilst individually are limited in power and bandwidth) can be clustered together with economic benefits. This multi-GPU environment can inspect more data and therefore log more information for investigators. Overall it offers easier maintenance and therefore can be kept up to date easier. Finally clustering many of these devices may provide corporations with a better understanding as to what is occurring on their networks at a cheaper on-going cost.

Keywords
Raspberry Pi, Network Forensics, Sensors, Law, Cluster, Corporate Challenges

INTRODUCTION
Modern society has become dependent on the Internet to such a degree that our love for data sees many people using the Internet at home, on the way to and from work, and at work itself.

Only last year Facebook advised it had surpassed one billion users, Apple announced it had 50 billion downloads from their app store, and Google stated they had half a billion active accounts (Abrahamsson, 2013)

As a result of Internet activity, a security expert, Marcus Ranum, first introduced the idea of network forensics in the 90's (Ranum, 1998). His idea was to have the ability to capture or record the network, and later analyse the data so as to discover the source of any security breach. But how much traffic needs to be recorded to be useful?

Historically it was easy for a network investigator to capture everything, store it all and quickly look through the data to find suspicious information and port usage. At the time, there were limited users and very separate protocols and almost no encryption used. However, as data usage has increased and a standardisation for the majority of applications to use the HTTP protocol for transmission, the ease for investigators to review relevant information has reduced and therefore increased the number of man-hours (Mazurczyk, 2014)

The growing use of cloud computing, encryption and the movement to Bring Your Own Devices (BYODs) has also brought with it a number of challenges to investigators. For example, how can an investigator find relevant information relating to an incident, if all of the data has been encrypted and transferred from a mobile device? If this data now resides outside the corporate network, in a foreign country, what legal capability does the company have to investigate what information has potentially fallen into the hands of competitors?

Botnets and the use of the dark web are additional areas of concern for corporations. These computers may be compromised internal computers, which have unknowingly joined a network and can be remotely controlled. Many of these networks use sophisticated encryption, require minimal amounts of network traffic and attempt to blend their activity in with normal traffic on the network. How can a modern network investigator track this activity and ensure it can be detected, stopped from further infestation, and prevented in the future?

With the multitude of viruses, worms and trojans present on the Internet, the cost to the organisation to keep the corporations infrastructure secure and to reduce possible infestation, far outweighs the cost associated should an

outbreak occur within the organisation (Nguyen, 2004). However, it is also easy to be of the mindset that a firewall or IDS is a catch all, leave and forget device, when in fact many constantly need updates and maintenance.

Network forensics is becoming increasingly important for these reasons, but as a result, intelligence also becomes a fundamental aspect. If we can ensure that sensors on the network are highly tuned and use the latest information from other corporations, there will be a higher chance that any indicators of attack are picked up and a lower chance of a "true-negative" result and filter out possible deceptive data (Gupta, 2003)

CURRENT CHALLENGES

Currently, many organisations take the same old approach installing Network Intrusion Detection Systems (NIDS): The server is installed at the egress and ingress points to the network (Bellovin, 2002). These links are typically the fastest links, which also carry the majority of the corporation's data. As a result, fast but expensive and power hungry machines are utilised in an attempt to keep up.

Unfortunately, monitoring every packet is virtually impossible for these individual devices without creating either a delay in traffic delivery or a bottleneck (Schaelicke, 2005). Additionally, should a Denial of Service (DoS) attack occur, the device attempting to monitor all traffic, could in fact be the weakest point in the network. This could subsequently cause a network outage as it attempts to carry out its duties.

Therefore NIDS typically only conducts dip samples of the traffic. The frequency of these samples is based on the load of the NIDS system, the speed of the processor, the network speed and the network utilisation. (Schaelicke, 2005)

Monitoring the network based on the egress and ingress points alone also does not sufficiently detect the possibility of an internal attack. This type of attack may never transmit via those data collection points. For this reason, monitoring should be conducted irrespective of the collection point.

Additionally, as many as many of these devices generate substantial heat, dedicated cooling and power filtration systems are required. This means all data must transit the same data links, to the same data centre where the NIDS devices are located, potentially causing locations for network congestion and single points of failure. This single device also causes issues for administrators attempting to keep the device updated. Also, having a single device does not allow any testing of new software.

Capturing relevant data is another challenge to the investigator, particularly when dealing with high volumes of data. In many instances the investigator unfortunately is frustrated the ultimate evidence is unobtainable – a full packet capture of everything that occurred. As a result, many investigations have been reliant on artifacts of communication, which do provide an insight into what has occurred, however come from a myriad of devices including routers, firewalls, proxy servers and access logs for services. This results in large log files and difficulty interpreting what has occurred (Bai, 2013).

The question therefore is how to capture the relevant amount of data without capturing too little or too much and how do we know what is relevant and what's irrelevant?

Ideally, the attack type will dictate how much information we may wish to capture. In the case of a Denial Of Service (DOS) attack do we really need to capture every single packet from every single host? Or is it better to detect the attack and protect assets? In the case of a trial, should an expert witness be called to court, the chances of the court discussing every single packet is therefore minimal. I suggest the court would be rather interested in the total number of packets and how this was detrimental to the company.

However, in the case unauthorised access of a computer (section 440A Criminal Code of Western Australia) (Government, 2003) times and dates of access and what they did, including all commands, maybe extremely relevant to prove beyond reasonable doubt, how the offender gained access and what they did once they were inside. Section 440A also requires that the system was a secure system and therefore does not have an easily guessed password (such as 'password') in order for successful prosecution.

An alternative solution would allow a cluster of computers, which don't care where the sample points are located. Effectively any computer can be linked into a cluster and purport to serve the requirements of the service previously discussed, however it's been found that companies are now consistently looking for more

energy-efficient servers, often using low-power CPUs (Cox, 2013) (Rohr, 2011) (Abrahamsson, 2013).

A MORE COST EFFECTIVE SOLUTION

Many of the problems discussed maybe resolved by a small device, which was traditionally marketed towards the hobbyist user.

The Raspberry Pi is a credit-sized computer, which contains a 700Mhz ARM processor and either 256 or 512mb of RAM. The devices can run on a 5V 1A DC current, making it possible to link multiple devices into a standard computer power supply (Kiepert, 2013). As there are no moving parts and a small footprint, the cooling requirements for these devices are significantly cheaper to that of commercial servers.

Whilst the Raspberry Pi may lack resources, such as the ability to use gigabit Ethernet connections (only having a 100Mps network interface) clustering multiple devices would still allow more data to be interpreted to that of a single NIDs system. Further development in the near future may also find the Raspberry Pi to include a gigabit or faster Ethernet port, similar to other products like the Banana Pi, along with more memory and processing power (Powell, 2014)

The University of Southampton recently commissioned a super computer comprising of 64 Raspberry Pi nodes; 64 processors, 1 TB of memory and cost under AUD $4500 (Cox, 2013). The devices were individually powered by mobile phone chargers and racked using just Lego blocks. The systems ran Raspbian Operating system under a Message Passing Interface (MPI) to distribute the workload amongst the 64 nodes.

Large multimode clusters using this device have been used in amazing technical breakthroughs including but not limited to Beowulf Clusters to tackle engineering challenges (Cox, 2013), Cloud based servers for the purpose of different services/applications (Kiepert, 2013) and clusters to search the skies for meteors (Norman, 2013).

Traditional servers contain large arrays of spinning disk platters and fast access RAM. These traditional spinning disk platters are known for their slow seek times whilst large volumes of RAM require vast amounts of power to be useful. The costs associated with purchasing these large servers can run into the tens of thousands of dollars per unit. A Raspberry Pi cluster on the other hand does not require spinning disks, large volumes of ram and is substantially cheaper than a traditional server (Cox, 2013).

Recent projects have also found easy and cheap ways in which the Raspberry Pi can be mounted in a fashion which maximizes space (Kiepert, 2013), similarly discussion has launched with the possibility of mounting a large number of Raspberry Pi computers into a single 1U rack mount case.

Projects have also found the maintenance of the Raspberry Pi cluster can be completed using a central repository or as simple as changing out an SD card (Abrahamsson, 2013) (Andersen, 2011). This in turn causes a saving in man-hours spent maintaining servers and large infrastructure.

Each Raspberry Pi device can also be removed from the cluster, updated and reconnected, this could not be possible with a single device. To save further time, a remote update command could also be used to update all nodes automatically.

The ability to remove a node from the cluster also affords the administrators the ability to test software updates to ensure there's no software problems or conflicts. This ability would not be possible with a single device, without downtime and loss of potential evidence.

Additionally networks with more sampling sample points raise the ability for the clusters to capture more information. The Raspberry Pi cluster would therefore be able to capture more of the conversation of an attack. Unfortunately as the Raspberry Pi is limited in the hard drive space, a disk node would be required for the purpose of a central repository for logging and data retention.

To prevent this disk node becoming a single point of failure, a multimode environment is recommended. This allows another disk node to take over, in the event of hardware failure. The linux Distributed Replicated Block Device (DRBD) project has matured to a standard, which has seen the software make its way into commercial and production environments (Ellenberg, 2007). This software is more than capable to allow the Raspberry Pi cluster to record relevant information as well as schedule downtime and maintenance on a disk node, should it

be required.

Finally the open source community has already developed many tools, which analyse multiple streams of data, graph as well as report on what happened, how it happened and for how long. As an example the logstash project has developed an open source tool, which quickly allows the investigator to parse logs, index, search and graph the results (Borouchaki, 2009).

In the case of a defacement of a website, the centralised logging server allows the corporation to report on typical network usage. This allows the investigator to create statistical reports, for court purposes, as to how many people saw the defacement and how much it cost the company in lost revenue.

A constant challenge to the security professional is to demonstrate to management how much money has been saved, by purchasing a security device. As the cluster has the ability to report on overall utilisation as well as possible attack vectors, a graphical representation could be incorporated which allows management to easily see thwarted attacks.

Graphical representations have also been discussed with regards to network and security administrators as an easier way to manage their networks and detect anomalies (Harrop, 2004). Should graphical representations be incorporated with the reporting mechanisms earlier discussed, it could easily reveal overall problems with the network, which are otherwise shadowed by complexity.

CONCLUSION

Network forensics is a complex task, which can be aided by a collection of simple, cheap and effective devices, however it is not without its challenges.

Each single device is somewhat slower than a standard desktop computer and extremely slow to that of a large industrial server or edge protection device currently utilised by corporations. The cluster of devices also requires regular updates to each of the cluster nodes to ensure they're aware of current heuristics and signatures, which may signify a possible attack.

As the Raspberry Pi is limited to the size of an SD card, it potentially lacks the storage capacity to capture large streams of data without filling to capacity. Network storage devices or disk nodes using open source, high availability software would be required to alleviate this problem.

However, despite these challenges, we have already seen many projects linking these devices together to form a cluster with substantial power savings and significant processing returns. The reduced carbon footprint, lower use of power and memory and ease of for updates makes these devices an attractive solution to corporations. It is only natural they find their way into greater applications, such as clusters discussed in this paper, as their software and hardware matures. Many projects have already addressed the issues of load balancing, logging and maintenance discussed.

Also there are other free solutions available in the open source community, which would address the issues regarding storage and reporting of attacks. Significant advantages exist to this cluster environment in comparison to a single NIDs or several expensive servers.

Ultimately the challenges are achievable and the device offers the ability to sample more data, at reduced running costs and initial outlay. It presents the network forensic investigator a chance to intelligently reconstruct the series of events, by capturing more data, accurately.

Finally and importantly in cases where a breach has occurred and where charges have been preferred, the investigator is afforded a better opportunity to discover what occurred. This ultimately can be communicated to the court, in absolute certainty and lead to a conviction.

REFERENCES

Abrahamsson, P. (2013). Affordable and Energy-Efficient Cloud Computing Clusters: The Bolzano Raspberry Pi Cloud Cluster Experiment. *CloudCom, IEEE, 2*(1), 170-172.

Andersen, D. G., Franklin, J., Kaminsky, M., Phanishayee, A., Tan, L., Vasudevan V. (2011). Fawn: fast array

of wimpy nodes. *ACM, 54(7)*(2011), 101-109. doi: 10.1145/1965724.1965747

Bai, J. (2013). Feasibility analysis of big log data real time search based on Hbase and ElasticSearch. *ICNC*, 1166 - 1170. doi: 10.1109/ICNC.2013.6818154

Bellovin, S. M., Cheswick, W. R. (2002). Network Firewalls. *IEEE C. M., 32*(9), 50-57. doi: 10.1109/35.312843

Borouchaki, H. (2009). XpoLog enhances logstash with Augmented Search. *Journal of Engineering (Atlanta, Ga), 64.*

Cox, S. J., Cox J.T., Boardman R.P., Johnston S.J., Scott M. (2013). Iridis-pi: a low-cost, compact demonstration cluster. *Cluster Computing, 17*(2), 349-358. doi: 10.1007/s10586-013-0282-7

Ellenberg, L. (2007). DRBD 8.0.x and beyond Shared-Disk semantics on a Shared-Nothing Cluster. *LinuxConf Europe*.

Government, W. A. (2003). Criminal Code Act Compilation Act. *State Law Publisher*. Gupta, N. (2003). Determining the effectiveness of deceptive honeynets.

Harrop, W., Armitage, G. (2004). Intuitive Real-Time Network Monitoring Using Visually Orthogonal 3D Metaphors. *ATNAC*.

Kiepert, J. (2013). Creating a Raspberry Pi-Based Beowulf Cluster

Mazurczyk, W., Szczypiorski K., Hui T. (2014). Network forensics and challenges for cybersecurity. *annals of telecommunications - annales des télécommunications, 69*(7-8), 345-346. doi: 10.1007/s12243-014-0434-7

Nguyen, J. (2004). The impact of Microsoft Windows infection vectors on IP network traffic patterns. *CAIA Technical Report 040804A*.

Norman, M. (2013). Meteor Raspberry Pi cluster to teach parallel computing. *NH&S*, 150.

Powell, M. (2014). Redesign for barebones Raspberry Pi computer. *ANM 2014, 1*(1)

Ranum, M. (1998). Experiences Benchmarking Intrusion Detection Systems. *Network Flight Recorder*

Rohr, D., Bach, M., Kretz, M., Lindenstruth V. (2011). Multi-GPU DGEMM and high performance linpack on highly energy-efficient clusters. *Micro, IEEE, 31*(5), 18-27. doi: 10.1109/MM.2011.66

Schaelicke, L., Freeland, J. C. (2005). Characterizing Sources and Remedies for Packet Loss in Network Intrusion Detection Systems. *IEEE WCS 2005, 1*(1), 188-196. doi: 10.1109/IISWC.2005.1526016