Intelligent Techniques for Effective Network Protocol Security Monitoring, Measurement and Prediction

Emmanuel Hooper
Harvard-MIT-Yale Scholar
Cambridge, MA, USA
Yale University
New Haven, CT, USA
and Information Security Group
University of London
Royal Holloway,
Egham, Surrey, TW20 OEX, UK.
President, EHSC, USA.
ehooper@aya.yale.edu

Abstract

The major problem is the absence of effective techniques for network protocol security monitoring, measurement and prediction. This is due to the emerging complex network protocols whose patterns are not readily determined by current tools and methods. Current tools and methods cannot handle the complex topology or patterns of traffic for accurate prediction of network protocol behavior for a wide range of time-scales. This includes the problem of effective data mining and characterization of network protocol topology structures, due to emerging global technologies and multiple protocols that interacting across different network protocol layers. The research develops new efficient techniques for network protocol characterization, monitoring and data measurement. This includes the development of new mechanisms, tools and methods for protocol measurements, characterization and accurate prediction of network protocol behavior in a wide range of network multiple-protocol environments that interact across different network protocol layers. The research presents innovative approaches for effective and efficient management, security, resilience, testing, analysis, design and implementation of network protocols in multiple network environments.

1. Introduction

The major problem is the absence of efficient techniques for network protocol security traffic, monitoring and data measurement. This is due to the emerging complex networks, protocols and traffic whose patterns are not readily determined by current tools and methods. Current tools and methods cannot handle the complex topology or patterns of traffic for accurate prediction of network protocol behavior for a wide range of time-scales. This includes the problem of effective data mining and characterization of network protocol topology structures, due to emerging global technologies and multiple protocols that interacting across different network protocol layers.

1.1 Current Approaches and Their Problems

The major problem with current state of art approaches of security controls is that they do not provide effective or efficient techniques for network protocol traffic characterization, monitoring and data measurement. The current techniques do not provide sufficient analysis of the traffic, due to the increasing variations in emerging complex networks, protocols and traffic. Current tools and methods do not provide effective topology for complex network protocol traffic patterns for accurate prediction of expected network protocol behavior in a wide range of time-scales. The current approaches collect and map data and traffic without effective data mining and characterization of network protocol topology structures of current global technologies and multiple protocols across different network protocol layers. The current methods for distinguishing between normal traffic and anomaly traffic lack effective algorithmic analysis and responses [1, 5, 14, 15]. This is exacerbated by the increase in network protocol traffic for both normal traffic and anomaly packets in critical infrastructures [5, 15, 17]. Furthermore, current techniques cannot analyze the vast subcategories of increasing packet types and rates on fast networks [3, 13, 15, 16, 19].

2 Network Protocol and Packet Analysis

The Network traffic alert verification and audit data analysis are used to analyze alerts but do not perform datamining on the data [2, 3, 9, 10, 23]. This involves alert aggregation and correlation [3] and analysis of alerts in multi-intrusion detection environments [2, 8, 20]. Other features include a formal data model for the correlation of traffic and alert [14], analysis of alerts using lightweight protocols [24] and alert management systems [23]. These, however, involve analysis of the alerts for intrusion attempts after they have already been alerted by the IDS. Since these methods depend to some extent on the alert data from the alerting systems, they are subject to similar problems of false positives that occur in the network alerting systems. Audit data analysis involves the evaluation of traffic from network protocols, applications, databases and security logging systems for anomaly detection [12, 13, 21, 22]. This includes analysis of audit and protocol data using artificial intelligence [22], identification of vulnerabilities in systems [12] and data collection for network protocols forensics [21]. However, similar to the problems in alert verification methods, audit data analyses are subject to the limitations of the network protocol environments protocols, applications and databases from which the audit data is collected. These limitations include the problems of false positives and false negatives of network security monitoring nodes and log hosts, malfunctioning servers and logging processes which interrupt logging, the limited sizes or log hosts, large volume of unfiltered audit data and minimum relevant logging. Furthermore, the absence of strategic log management and archiving result in inefficient timely response to attacks in fast gigabyte environments. These are exacerbated by the vulnerabilities in protocols, application and databases and their interface connections to various internal network zones and the DMZ. Other network traffic data analysis tools such as honeypots and honeynets collect data sources consisting of suspicious traffic with some interaction and provide limited interaction with the intruder [4]. However, the logging requires more instructions and virtual machines require more time to execute the instructions and consist of security vulnerabilities [6, 7, 11, 18]. Furthermore, Honeynets do not interact

or send any feedback to the network or similar alerting systems and have significant costs depending on the implementation of the network environment and platform.

3. New Intelligent Techniques for Network Protocol Security, Monitoring, Measurement and Accurate Prediction

The new approach develops new efficient techniques for network protocol security monitoring data measurement and accurate prediction of network protocol behavior in a wide range of time-scales and multiple protocols that interact across different network protocol layers. It also develops effective techniques for network protocol simulation models and algorithms for large-scale heterogeneous networks. First, the research develops efficient techniques for network protocol traffic characterization, monitoring and data measurement. Secondly, the research develops new techniques for analyzing network protocol data accuracy in large-scale sensor networks. Thirdly, the research develops effective techniques for effective simulation algorithms for large-scale, heterogeneous networks for identification and extraction of meaningful network protocol patterns for network protocol simulation toward accurate depiction of network protocol behaviors.

We analyze patterns of network protocol traffic for characterization, monitoring and data measurement for accurate prediction of network protocol behavior in a wide range of time-scales and technologies that use multiple protocols to interact across different network protocol layers. We examine effective design networks and data storage for processing for accuracy in large-scale networks. This includes the development effective strategies both real-time and off-line and effective algorithms for characterization and prediction of network protocol behavior and anomaly network protocol traffic patterns on large-scale heterogeneous networks for effective mitigation strategies.

4. Network Protocol Architecture

The architecture for analyzing network protocol patterns and detecting and predicting complex network protocol behaviors for developing datamining algorithms involving packet data analysis is shown in Figure 1. This involves packet capturing and filtering of relevant attributes from multiple network protocol packet types and protocol analysis and event logs to a central database for analysis of potential intrusion. A combination of packet data analysis and statistical anomaly can be used to predict and mitigate impact on network protocol behavior such as network attacks.

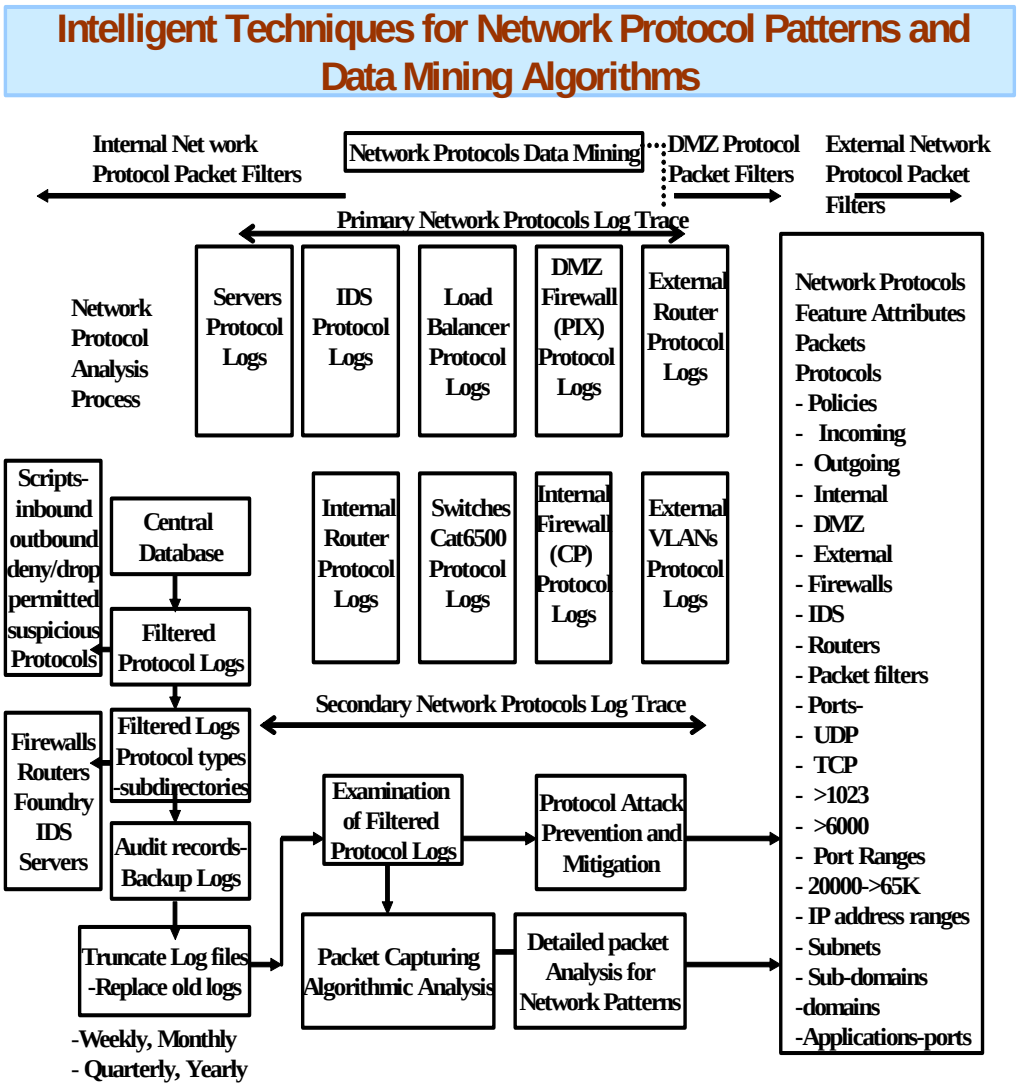

Figure 1. Architecture for monitoring, measurement and prediction of network protocol patterns

5. Research Method

The research methods involves analysis of the patterns of network protocol traffic for characterization, monitoring and data measurement for accurate prediction of network protocol behavior in a wide range of time-scales and technologies that use multiple protocols to interact across different network protocol layers. We log, trace, aggregate, correlate and perform data mining on critical large-scale networks and data storage including for accurate prediction of packet types, severity of impact for containment of adverse impact. This includes datamining analysis of critical network protocol topology and traffic for development of accurate prediction algorithms for characterization and prediction of network protocol behavior and anomaly network protocol traffic patterns on large-scale heterogeneous networks for effective mitigation strategies

The experiments involving the adaptive firewall packet filters in responses from the NQCs were conducted as follows. The method involves the diversion of suspected network attacks to the quarantined channel zones. This is followed by sending responses to the suspicious packets, which appear as valid return packets to the potential attacker. This results in further packets from the attacker which if they persist is directed to subsequent zones for additional responses.

6. Steps in Experimental Methods: Data Mining Analysis

6.1 Network Protocol Traffic Pattern Subcategories

The datamining algorithms including rule induction and program were written to examine the prediction and accuracies of the subcategories of network protocol packets. The following example shows an algorithm for network traffic behavior patterns that predicts pattern subcategory. This rule induction of packet conditional feature attributes are used for successive predictions of subcategories of network protocol packet types for accurate mitigation of risks.

6.2 Model Conditional Program for Developing Algorithm for Network Protocol Pattern Analysis

The following code segment is a sample of the model conditional program for developing the network pattern protocol analysis algorithm.

Conditional Program for Network Protocol Algorithm

```
If ProtocolType = Exploit And ApplicationProtocol
        = radius Or ApplicationProtocol = rsh Or
        ApplicationProtocol = ftp Or ApplicationProtocol
        = netbios And ResultStatus = Successful And
        Category = Exploit And DetectionMechanism
        = signature Then protocol subcategory-type
        = bufferoverflow
ElseIf ProtocolType = Exploit And ApplicationProtocol
        OR telnet And Category = Exploit And
        SourcePort = undetermined And
        ApplicationProtocol = undetermined And
        ResultStatus = Suspicious
```

And Category = V olumeDoS And
DetectionMechanism = statisticalanomaly Then
protocol subcategory-type = statisticaldeviation.
End If

7. Results

The research results for network protocol pattern prediction accuracies using various data mining techniques are shown in Tables 1 to 3. The overall accuracies were 99.83% for training data and 99.64% for test data selected at random from approximately 200 Gigabytes of network traffic in a large-scale commercial environment. Approximately 40% of the data was used for training and 60% was used for the test data. These accuracies pertain to rule induction datamining of the conditional attributes of traffic features for attack patterns in the database and distinguish between normal connections and anomaly attack traffic patterns in real-time. The final result for subcategories of network protocols is 99.95%. See Table 3.

Table 1. Results: Network Protocol Analysis Training Dataset – Normal/Anomaly Confusion Matrix Accuracy Summary

Class Target	Positives	Negatives	Test Accuracy
Normal	**20,100**	**54**	**99.73%**
DCERPC DOS	4,000	25	**99.38%**
DDOS	3,500	11	**99.69%**
ICMP Volume Too High	2,500	0	**100.00%**
HTTP Path Too Long	1,000	0	**100.00%**
MSSQL Buffer Overflow	520	0	**100.00%**
SMTP Worm	450	0	**100.00%**
Total:	**Actual Positives**	**Actual Negatives**	**Accuracy**
	11,970	**36**	**99.70%**

Table 2. Results: Network Protocol Analysis Test Dataset – Normal/Anomaly Confusion Matrix Accuracy Summary

Class Target	Positives	Negatives	Training Accuracy
Normal	**32,500**	**160**	**99.51%**
DCERPC DOS	6,500	62	**99.06%**
DDOS	5,500	20	**99.64%**
ICMP Volume Too High	3,850	0	**100.00%**
HTTP Path Too Long	1,450	0	**100.00%**
MSSQL Buffer Overflow	850	0	**100.00%**
SMTP Worm	692	0	**100.00%**
Total:	**Actual Positives**	**Actual Negatives**	**Accuracy**
	18,482	**82**	**99.57%**

Table 3. Results: Network Protocol Pattern Analysis Test Dataset – Normal/Anomaly Confusion Matrix Accuracy Summary

Class Target	Positives	Negatives	Test Accuracy
Normal	80,124	38	**99.95%**
Protocol Statistical - Deviation	3,456	0	**100.00%**
Protocol Anomaly	1,427	2	**99.86%**
Protocol Malware Code Exploit	4589	4	**99.91%**
Protocol Complex Attack	1452	0	**100.00%**
Total :	**Actual Positives**	**Actual Negatives**	**Accuracy**
	91,048	**44**	**99.95%**

8. Algorithm for Accurate Prediction of Multiple Network Protocol Subcategories

The algorithm for protocol subcategory for the efficient prediction network of patterns is defined as follows.

1. Select each Network Protocol dataset
2. Select one of the data training samples, x of the dataset.
3. Select and define the relevant feature attributes a, of datasets = D (Ca), such that: Caj e D, a = 1,., n, and j = 1,., N, where C is a class member, denoted by e of dataset D, for all class types C1 to Cn, each with attribute values j, from 1 to N.
4. Define the network traffic class type and data structure within each dataset for network protocol event packet class.
5. Apply Algorithms e.g. Rule Induction for network protocol attribute features to obtain conditional rulesets.
6. Filter for 100% (maximum) support of rulesets.
7. Induct conditional rules using filtered rulesets for all subcategories, protocol subcategory S1 protocol subcategory SN

A. For all classes, Class C1 Class CN
   i. Select features common to class.
   ii. Select features different from class using statistical analysis for Class types 1 to N for all variant features within class type.

B. Repeat Step A for Next Class
C. Repeat Steps A and B for the Next protocol subcategory

8. Select sample intrusion test data
9. Repeat Steps 1 to 7 for test data
10. Repeat Steps 1 to 9 and aggregate results for all network protocol types

9. Discussion

Real datasets provide realistic, accurate and comprehensive datasets for network protocol security pattern detection in real application and network protocol infrastructure environments. In addition, the attribute features of protocol packet logs from network routers, switches, servers, load balancers, IDS and firewalls including events, time, destination and source IP addresses and ports, were correlated to analyze the patterns of network behavior across various network segments, subnets and zones. Furthermore, the complex attacks at the TCP level were captured through the status of the TCP flags in each session. These combined datasets provided cumulative validation of the framework and algorithmic process for a greater accuracy in prediction of network behavior patterns.

10. Conclusion

The intelligent techniques are applicable for network protocol security, monitoring, measurement, and accurate prediction. The research results in the development of effective solutions and techniques for network protocol characterization, monitoring and data measurement. This includes the development of new mechanisms and methods for protocol subcategory measurements and accurate prediction of network behavior in multiple network protocols. Furthermore, the research results in the effective design of network protocol patterns, packet analysis and accuracy in large-scale sensor networks. This includes the

strategic intelligent development of effective strategies intelligent techniques for network protocols and effective algorithms for network protocol monitoring, measurement, and accurate prediction. Furthermore, it provides effective means for efficient management, security, resilience, testing, analysis, design and implementation of network protocols in multiple network environments.

11. References

[1] T. Bowen, D. Chee, and M. Segal. "Building survivable systems: An integrated approach based on intrusion detection and damage containment." In IEEE Proceedings of the DARPA Information Survivability Conference and Exposition, volume II of II, IEEE Computer Society Press, 2000, pages 84–999.

[2] F. Cuppens. Managing alerts in multi-intrusion detection environment. In *Proceedings 17th Annual Computer Security Applications Conference*, pages 22–31, New Orleans, 2001.

[3] H. Debar and A. Wespi. Aggregation and correlation of intrusion-detection alerts. In *Recent Advances in Intrusion Dteection (RAID2001)*, volume 2212 of Lecture Notes in Computer Science, pages 85–103. Springer-Verlag, Berlin, 2001.

[4] J. B. Grizzard, C. R. Simpson, Jr., S. Krasser, H. L. Owen, and G. F. Riley. Flow based observations from NETI@home and Honeynet data. In *Proceedings of the 2005 IEEE Workshop on Information Assurance and Security*, United States Military Academy, West Point, NY, USA, June 15–17 2005. IEEE Computer Society Press. http://www.megasecurity.org/papers/defeatinghpsiaw05.pdf.

[5] G. Helmer, J. Wong, V. Honavar, and L. Miller. Intelligent agents for intrusion detection. In *Proceedings of the 2003 IEEE Information Technology Conference*, pages 121–124, Syracuse, NY, USA, September 1998. IEEE Computer Society Press.

[6] T. Holz and F. Raynal. Detecting honeypots and other suspicious environments. In *Proceedings of the Sixth IEEE Workshop on Information Assurance and Security*, United States Military Academy, West Point, NY, USA, June 15–17 2005. IEEE Computer Society Press.

[7] T. R. Jackson, J. G. Levine, J. B. Grizzard, and O. H. L. An investigation of a compromised host on a honey net being used to increase the security of a large enterprise network. In *Proceedings of the 5th Annual Information Assurance Workshop*, pages 9–15, West Point, NY, USA, 2004. IEEE Computer Society Press.

[8] C. Kruegel, W. Robertson, and G. Vigna. Using alert verification to identify successful intrusion attempts. *Practice in Information Processing and Communication (PIK)*, 27(4), October 2004. [9] C. Kruegel, T. Toth, and E. Kirda. A security policy reinforcement tool for large networks. In *Proceedings of IFIP Conference on Advances in Network and Distributed Systems Security*, Boston, MA, USA, November 2002. Kluwer Academic Publishers.

[10] C. Kruegel, T. Toth, and E. Kirda. Decentralized event correlation for intrusion detection. In *Proceedings of International Conference on information Security and cryptology*, volume Lecture Notes in Computer Science 2288, Berlin, Germany, December 2006. Springer-Verlag, Berlin.

[11] J. Levine, R. La Bella, H. Owen, D. Contis, and B. Culver. The use of honeypots to detect exploited systems across large enterprise networks. In *Proceedings of the 2003 IEEE Workshop on Information Assurance*. IEEE Computer Society Press, 2003.

[12] R. P. Lippmann, S. E. Webster, and D. Stetson. The effect of identifying vulnerabilities and patching software on the utility of network intrusion detection. *Computer Networks: The International Journal of Computer and Telecommunications Networking*, 3949 of Lecture Notes in Computer Science:307–326, 2002.

[13] M. V. Mahoney and P. K. Chan. An analysis of the 1999 DARPA Lincoln Laboratory evaluation data for network anomaly detection. In *Recent Advances in Intrusion Detection (RAID2003)*, volume 2820 of Lecture Notes in Computer Science, pages 220–237. Springer-Verlag, Berlin, 2003.

[14] B. Morin, L. Me', H. Debar, and M. Ducasse. M2D2: A formal data model for IDS alert correlation. In *Recent Advances in Intrusion Detection (RAID2002)*, volume 2515 of Lecture Notes in Computer Science, pages 115–137, Zurich, Switzerland, 16–18, October 2002. Springer-Verlag, Berlin.

[15] Network Associates. McAfee Intrushield IDS: 4000 Series, 2007. Santa Clara, CA, USA.

[16] V. Paxson. Bro: A system for detecting network intruders in real-time. *Computer Networks*, 31(23– 24):2435–2463, 1999.

[17] L. Portnoy, E. Eskin, and S. Solfo. Intrusion detection with unlabelled data using clustering. In *Proceedings of ACM CSS Workshop on Data Mining Applied to Security (DMSA-2001)*, pages 76–105, 2001.

[18] N. Provos. A virtual Honeypot framework. CITI Technical Report 03-1, Center for Information Technology Integration, University of Michigan, Ann Arbor, MI, 2003.

[19] T. H. Ptacek and T. N. Newsham. Insertion, evasion and denial of service: Eluding network intrusion detection. Technical Report, Secure Networks (McAfee) Inc., Santa Clara, CA, USA, January 1998. http://citeseer.ist.psu.edu/ptacek98insertion.html.

[20] F. Valeur, G. Vigna, C. Kruegel, and R. Kemmerer. A comprehensive approach to intrusion detection alert correlation. *IEEE Transactions on Dependable and Secure Computing*, 1(3):146–169, July 2004.
[21] W. Wang and T. E. Daniels. Building evidence graphs for network forensics analysis. In *21st Annual Computer Security Applications Conference (ACSAC'05)*, pages 254–266, Tucson, AZ, USA, December 2005. IEEE Computer Society Press.
[22] W. R. Weiss and A. Baur. Analysis of audit and protocol data using methods from artificial intelligence. In *Proceedings of the 13th National Computer Security Conference*, pages 109–114, Washington, D.C., USA, October 1990.
[23] J. Yu, Y. V. R. Reddy, S. Selliah, S. Kankanahalli, S. Reddy, and V. Bharadwaj. TRINETR: An intrusion detection alert management system. In *13th IEEE International Workshops on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE'04)*, pages 235–240. IEEE Computer Society Press, December 2004.
[24] J. Zhou, A. J. Carlson, andM. Bishop. Intrusion alerts using lightweight protocol analysis. In *Proceedings of the 21st Annual Computer Security Applications Conference (ACSAC'05)*, pages 117–126. IEEE Computer Society Press, 2005.

Authors

Emmanuel Hooper

Dr. Emmanuel Hooper earned a record of 3 PhDs within 5 years. He earned a PhD at the University of London, Royal Holloway, Information Security Group, UK, 2007; a PhD in Computing Sciences from the University of East Anglia, UK, 2006; and PhD in Historical Statistical research from the University of Birmingham, UK, 2005. He holds a BSEE from Portsmouth University, UK, multiple MA degrees from various universities including Yale University, USA, including graduate and postdoctoral studies at Oxford, Cambridge, Harvard, MIT and Yale Universities. He has 28 years experience in infrastructure security and is an adjunct faculty member at the University of California, Riverside, USA. He is a member of various organizations including IEEE, a researcher and consultant in security for various major US and UK companies and President of EHSC/CISO in Palo Alto, California, USA and a Harvard-MIT-Yale Cyber Scholar.