Bidirectional Flow Measurement, IPFIX, and Security Analysis

Elisa Boschi
Hitachi Europe SAS
Sophia Antipolis Laboratory
Valbonne, France

Brian Trammell
CERT Network Situational Awareness Group
Carnegie Mellon University
Pittsburgh, PA, USA

Abstract

This paper describes the addition of bidirectional flow export to the IPFIX protocol, and the impact of this effort on security-related flow analysis. Along the way, it examines the application of bidirectional flow measurement to common security analysis tasks and the positive impact the adoption of IPFIX as a common interchange format could, and will, have on the community using flow measurement for security purposes.

1 Introduction

Bidirectional flow (biflow) collection and analysis are broadly applicable to security-related network measurement and flow analysis tasks. The IETF's IPFIX working group has defined a new protocol [5] for flow information export, based on Cisco's NetFlow V9, which provides a flexible representation for a wide variety of flow data; however, the original standard has no native support for biflow measurement. Therefore, adding biflow support to IPFIX would seem to be a next logical step.

This paper describes the addition of biflow export to IPFIX, an effort that began at FloCon 2005. The attendees of that workshop identified biflow export as a significant missing feature in the IPFIX flow export protocol. The effort that followed is an excellent example of cooperation between the standards community and the user communities involved in security-oriented network flow analysis.

In section 2, we review the motivation behind biflow measurement and examine its uses. We describe the IPFIX protocol and its applicability to security-oriented network flow analysis in section 3. We then bring the two together, and demonstrate how biflow export has been added to IPFIX in section 4. We conclude with remarks on this effort and its future in section 5.

2 Bidirectional Flow Measurement

Bidirectional flow (or biflow) measurement refers to the association of information about both directions of an internet traffic flow in the collected data. This association provides additional data that can be useful for common analysis tasks. Trivial examples of biflow applications include initial round trip time estimation, the detection of connection establishment or other transactions for the purposes of incident detection and response, and the separation of unanswered traffic for scan detection purposes.

The network measurement and network security research literature do not have much to say about biflow measurement. This is largely explained by the fact that most research studies collecting omnidirectional packet data for the purposes of discovering new aspects of the structure of the Internet, the protocols that define it, or the endpoint implementations that comprise it, generally use full or sampled packet trace or payload data without reducing those packets into flows. When the packet data is available and of a manageable size, this reduction is unnecessary. A bidirectional data model often lurks at the core of these efforts, as in [10] and [18]. Indeed, bidirectional flow state is essential for any network measurement system that models the flow state of end systems, as the Argus [13] flow meter and the ubiquitous Bro [12] and Snort [15] network intrusion detection systems do.

That said, the concept of bidirectional flow data is not a new one. The RTFM [3] flow metering protocol supported bidirectional counters, as does Argus, which supports its own metering protocol. Biflow measurement does place one restriction on the design of a metering system: either meters must be placed such that each meter observes packets in both directions of the flow and associates the halves of each biflow at observation time, or the collectors of unidirectional flow data from the meters must centralize the data for matching purposes. The latter design requires $O(n^2)$ comparisons to match n simultaneously active observed connections. Clearly, the

earlier in the measurement process that the biflow association can be made, the lower the total resource requirements across the system. Any meter deployed at a single network perimeter or an Internet access point in the absence of asymmetric routing is well-positioned for biflow measurement.

Unidirectional flow (or uniflow) data is the still the type with which the security community is most familiar, owing to the ubiquity of the unidirectional Cisco NetFlow [4], its widespread existing deployment as a billing solution by many network service providers and enterprises, and the ecosystem of tools and techniques that have grown around it [6, 7]. However, when the metering system architecture supports it, biflow measurement is beneficial to the general task of flow analysis for research and operational security purposes.

3 IPFIX: A Standard Flow Export Protocol

The IPFIX protocol is the upcoming IETF standard for IP flow information export. It specifies how IP traffic flow information can be transmitted over the network from routers, measurement probes or other devices to a collector by providing a common representation of flow data and a standard means of communicating them. In this section we highlight the benefits such a standard format will bring to the security community and give an overview of the protocol itself.

3.1 Protocol Description

IPFIX is a unidirectional, template-based data transport protocol that provides flexible flow selection; a flow can be defined by an arbitrary number of packet fields (as opposed for instance to the five-tuple in NetFlow Version 5 [4]), which compose the flow key. The properties used to select flows, and therefore the flow key, can vary depending on the target application.

Flow information is exported using flow data records and template records. Templates contain {type, length} pairs specifying which {value} fields are present in data records conforming to the template, giving great flexibility as to what data is transmitted. Since templates are sent very infrequently compared to data records, this mechanism is much more efficient than formats that add descriptive information to each data record (e.g., XML). Different data record formats may be transmitted simply by sending new templates specifying the {type, length} pairs for the new data format.

IPFIX defines a set of information elements for describing flows, which provide the necessary {type} information for templates. This information model [14] makes IPFIX an open, interoperable flow information export protocol. Each information element in this model is identified within the protocol by a well-known number. The information model contains information elements for most common flow export tasks: the five-tuple, other IP and TCP header fields, fields describing packet treatment, a variety of counters, and timestamps with precision from seconds down to nanoseconds. The protocol also allows the definition of "private" information elements, to allow the extension of this information model for purposes private to a single vendor or enterprise without going through the full process required to create a new public information element.

3.2 Applying IPFIX to Security Analysis

Adoption of IPFIX for exporting and representing flow data will provide many benefits to the security-oriented flow analysis community.

First of all, as the emerging standard for flow export, IPFIX provides the advantages of a standard solution. IPFIX-compliant meters, collectors, and intermediate processes will interoperate easily. Broad support for the standard will lead to widespread implementations from a variety of vendors, which will subsequently lead to broad deployment in the field. IPFIX can then be used as a common input interface for analysis tasks, which will consequently reduce the need for ad-hoc solutions.

Moreover, IPFIX's templated data format provides significant flexibility. Tools can interoperate on common information while keeping their own internal data models and private information in each record. The template mechanism also allows easy addition of information to record formats, adding to the flexibility of tools to adapt to new requirements while mainitaining forward and backward compatibility.

4 Bidirectional Flow Export using IPFIX

The broad applications of biflow data to various security-related network measurement efforts, and the usefulness of a standard interchange format to support implementation efficiency and cross-domain information sharing have proved to be strong motivations for an IPFIX extension supporting biflow export.

The most recent result of this ongoing effort is an Internet-Draft, "Bidirectional Flow Export using IPFIX" [16], which has been accepted as a work item for the IPFIX Working Group and is slated to become and Informational RFC [1].

As defined in this draft and used throughout this document, a biflow is a flow composed of a number of packets sent in both directions between two endpoints, where a flow is simply a collection of packets sharing common

packet header fields and other characteristics (cf., the IPFIX definition of "IP Traffic Flow" [5]).

There are compelling reasons to export biflows as single entities within IPFIX. The usefulness of bidirectional association in flow data has already been established; in addition, exporting biflows as single entities can result in improved export efficiency by eliminating duplicate flow key data from the IPFIX message stream.

Biflow export with the IPFIX protocol, as proposed in the draft, will use a single flow record to represent each biflow. Single-record export does introduce some additional semantic considerations. When handling uniflows, the semantics of "source" and "destination" information elements are clearly defined by the semantics of the underlying packet header data. When grouping a biflow into a single data record, the definitions of "source" and "destination" become less clear.

We resolve this difficulty by defining the source and destination addresses of the flow as the source and destination addresses of the packet initiating the flow, respectively. The biflow is then defined to have two directions. The "forward" direction contains packets sent from the flow source to the flow destination, and the "reverse" direction contains packets sent the other way. Each biflow then has two sets of non-key fields, one for each of these directions.

Choosing direction by biflow initiator can be roughly approximated by a metering process by simply assuming the first packet seen in a given biflow is the packet initiating the flow. Some metering technologies may improve upon this method using some knowledge of the transport or application protocols (e.g., TCP flags, DNS question/answer counts) to better approximate the flow-initiating packet.

Creating a "reverse" information element counterpart to each presently defined "forward" information element will cover most or all of the information model, since only certain identifiers and metering and export process properties (cf. the IPFIX Information Model [14]) are not subject to reversal. Single-record biflow export will require a great number of new reverse information elements. The IPFIX Information Model has more than adequate number space for official information element expansion. However, the additional reverse information elements are not so much a discrete list of new information elements as a new dimension in the information model.

This new reverse information element number space can be created by leveraging the vendor-specific information element feature of the IPFIX protocol. This feature allows an entity other than IANA to define information elements by scoping them to an IANA-assigned Private Enterprise Number (PEN) [8]. This draft proposes defining a special PEN to signify "IPFIX Reverse Information Element". This reverse PEN serves as a "reverse direction flag"in the template; each information element number within this PEN space is assigned to the reverse counterpart of the corresponding IANA-assigned public information element number.

5 Conclusions and Future Work

IPFIX defines a message format, information model, and wire protocol for the collection of network flow data. All of the Internet-Drafts defining the protocol are near the end of the standards process and should be completed by the end of 2006. IPFIX implementations will begin to supplant the various NetFlow versions in use to become the standard for flow export over the coming years; the security-oriented flow analysis community can leverage this new standard to provide a basis for interoperable implementations of the entire flow collection and analysis process, and to facilitate the sharing of data among tool chains and across administrative domain boundaries.

As biflow measurement is beneficial for many common flow analysis tasks, and the IPFIX protocol as it stood in 2005 had no native support for representing biflows, the authors undertook an effort to add biflow export to the protocol, in order to facilitate biflow collection all the way back down the collection stack to the flow metering process. This effort has been successful; it is expected the addition of biflow export support to IPFIX will be published as an Informational RFC in early 2007 [9]. This will represent an approximately eighteen-month lifecycle from conception at FloCon 2005, at the suggestion of the assembled workshop, to completion.

This provides an excellent example of how the IETF standards community and various protocol user communities – in this case, the security-oriented flow analysis community represented at FloCon – can interact and cooperate in order to further the application of standards and the goals of the communities in question. Standards are good for this community: by adopting a common interchange format, flow meters and collectors can interoperate, easing the deployment of flow measurement both within and across organizational boundaries. Likewise, this community has been good for the standards process: without the effort described here, one fewer potential user community would be well-served by the IPFIX standard. The authors look forward to being part of the continued relationship between these two communities.

As such, the authors are also involved in another enhancement to IPFIX that we hope will benefit the security-oriented flow analysis community. The IPFIX protocol, as noted, is designed for the export of flow information from metering processes, through any number of optional intermediate processes, toward a fi-

nal collecting process for storage, reporting and analysis. This one-way data flow, required by IPFIX's nature as a unidirectional message-stream oriented protocol, is well suited for the initial collection of flow data. It would not seem to apply as readily to the generally more fluid, document-oriented workflows used by analysis tool suites such as the NCSA visualization tools [1] or the SiLK suite [7, 11].

However, as also noted, at the core of the IPFIX protocol lies a simple, efficient, flexible message format on which a common document format could easily be based. Such a document storage format could unify the information model and data representation used by the variety of analysis tool chains from metering and collection through storage and final analysis. To that end, the authors are presently working on a new Internet-Draft [17] to define the requirements for and the technical details of this format.

6 Acknowledgments

Special thanks to the attendees of FloCon for bringing the biflow issue to the attention of the IETF IPFIX working group, and to the members of the IETF IPFIX working group for attending to it. Thanks also to Lutz Mark, Juergen Quittek, Andrew Johnson, Paul Aitken, Benoit Claise, and Carsten Schmoll for their contributions and comments to the biflow export draft itself.

References

[1] BEARAVOLU, R., LAKKARAJU, K., AND YURCIK, W. NVisionIP: An Animated State Analysis Tool for Visualizing NetFlows. In *FloCon 2005: Proceedings* (Pittsburgh, Pennsylvania, USA, September 2005), CERT Network Situational Awareness Group.

[2] BRADNER, S. The Internet Standards Process – Revision 3. RFC 2026 (Best Current Practice), Oct. 1996. Updated by RFCs 3667, 3668, 3932, 3979, 3978.

[3] BROWNLEE, N., MILLS, C., AND RUTH, G. Traffic Flow Measurement: Architecture. RFC 2722 (Informational), Oct. 1999.

[4] CISCO SYSTEMS. Cisco IOS Netflow. `http://www.cisco.com/en/US/products/ps6601/products_ios_protocol_group_home.html`, 2006. Last Visited: 28 June 2006.

[5] CLAISE, B. IPFIX Protocol Specification, June 2006. INTERNET-DRAFT draft-ietf-ipfix-proto-22 (work in progress).

[6] FULLMER, M., AND ROMIG, S. The OSU Flow-tools Package and Cisco Netflow Logs. In *Proceedings of the 14th Systems Administration Conference (LISA 2000)* (New Orleans, Louisiana, USA, December 2000), USENIX Organization, pp. 291 – 303.

[7] GATES, C., COLLINS, M., DUGGAN, M., KOMPANEK, A., AND THOMAS, M. More NetFlow Tools: For Performance and Security. In *Proceedings of the 18th Large Installation Systems Administration Conference (LISA 2004)* (Atlanta, Georgia, USA, November 2004), USENIX Organization, pp. 121–132.

[8] INTERNET ASSIGNED NUMBERS AUTHORITY. IANA Private Enterprise Number registry. `http://www.iana.org/assignments/enterprise-numbers`.

[9] INTERNET ENGINEERING TASK FORCE. IP Flow Information Export (ipfix) Charter. `http://www.ietf.org/html.charters/ipfix-charter.html`, June 2006. Last Visited: 29 June 2006.

[10] KOMPELLA, R. R., SINGH, S., AND VARGHESE, G. On Scalable Attack Detection in the Network. In *Proceedings of the Internet Measurement Conference 2004* (Taormina, Sicily, Italy, October 2004), Association for Computing Machinery, pp. 187 – 200.

[11] MCNUTT, J. R: A Proposed Analysis and Visualization Environment for Network Security Data. In *FloCon 2005: Proceedings* (Pittsburgh, Pennsylvania, USA, September 2005), CERT Network Situational Awareness Group.

[12] PAXSON, V. Bro: A System for Detecting Network Intruders in Real-Time. *Computer Networks 31*, 23 – 24 (December 1999), 2435 – 2463.

[13] QOSIENT, LLC. argus: network Audit Record Generation and Utilization System. `http://www.qosient.com/argus/`, 2004. Last Visited: 9 May 2006.

[14] QUITTEK, J., BRYANT, S., CLAISE, B., AND MEYER, J. Information Model for IP Flow Information Export, June 2006. INTERNET-DRAFT draft-ietf-ipfix-info-12 (work in progress).

[15] ROESCH, M. Snort: Lightweight Intrusion Detection for Networks. In *Proceedings of LISA '99: 13th Systems Administration Conference* (Seattle, Washington, USA, November 1999), USENIX Organization, pp. 229 – 238.

[16] TRAMMELL, B., AND BOSCHI, E. Bidirectional Flow Export using IPFIX, June 2006. INTERNET-DRAFT draft-trammell-ipfix-biflow-02 (work in progress).

[17] TRAMMELL, B., BOSCHI, E., MARK, L., AND ZSEBY, T. An IPFIX-based File Format, June 2006. INTERNET-DRAFT draft-trammell-ipfix-file-01 (work in progress).

[18] WATSON, D., SMART, M., MALAN, G. R., AND JAHANIAN, F. Protocol Scrubbing: Network Security Through Transparent Flow Modification. *IEEE/ACM Transactions on Networking 12*, 2 (April 2004), 261 – 273.

Notes

[1]The Internet Standards Process, which defines the IETF-specific terminology that appears in this section and this document in general, is described in RFC 2026 [2]