A More Practical Approach for Novel Hybrid Schemes of Packet Marking and Logging for IP Traceback

V S R S Phani Kumar, P. Karunakar Reddy, S. Suresh

Dept. of CSE, Sree Rama Institute of Technology & Science, SKuppenakuntla, Penuballi, Khammam, AP, India

Abstract - Computer network attacks are on the increase and are more sophisticated in today's network environment than ever before. Tracing DoS attacks that employ source address spoofing is an important and challenging problem. Traditional traceback schemes provide spoofed packets traceback capability either by augmenting the packets with partial path information (i.e., packet marking), or by storing packet digests or signatures at intermediate routers (i.e., packet logging). Such approaches require either a large number of attack packets to be collected by the victim to infer the paths (packet marking), or a significant amount of resources to be reserved at intermediate routers (packet logging). We adopt a hybrid traceback approach in which packet marking and packet logging are integrated in a novel manner, so as to achieve the best of both worlds, that is, to achieve small number of attack packets to conduct the traceback process and small amount of resources to be allocated at intermediate routers for packet logging purposes. Based on this notion, two novel traceback schemes are presented. The first scheme, called Distributed Link-List Traceback (DLLT), is based on the idea of preserving the marking information at intermediate routers in such a way that it can be collected using a link list based approach. The second scheme, called Probabilistic Pipelined Packet Marking (PPPM), employs the concept of a "pipeline" for propagating marking information from one marking router to another so that it eventually reaches the destination. We evaluate the effectiveness of the proposed schemes against various performance metrics through a combination of analytical and simulation studies.

I. Introduction

Denial-of-service (DoS) attacks have been threatening the utility of the Internet severely [1]. In 2002, a coordinated attack on the Internet name service infrastructure showed the possibility and potential impact of such dedicated attacks [2]. More recently, it was reported that DoS attacks have been used as a means of extortion and become the subject of lawsuits [3]. Defending against DoS attacks requires not only measures for mitigating the effects of the attacks but also mechanisms for identifying the entities accountable for such attacks. Denial of Service (DoS) attacks [11], [12] and a more complicated version known as Distributed DoS (DDoS) are the most common to take advantage of source address spoofing. These attacks deny regular Internet services from being accessed by legitimate users either by blocking service completely or by disturbing it such that users become not interested in the service anymore (for example causing significant delay in accessing an airline reservation website). In such attacks, the main objective is to overpower the victim while concealing attacker's identity. Today's Internet has witnessed several incidents that confirm the devastating effect of such attacks. For example, in October 2002, eight out of the thirteen root DNS servers were brought down as a result of severe flooding denial of service attack [11]. IP traceback is the process of identifying the actual source(s) of attack packets. This has the benefit of holding attackers accountable for abusing the Internet. Also, it helps in mitigating DoS attacks either by isolating the identified attack sources, or by filtering attack packets far away from the victim as proposed in the IP traceback-based intelligent packet filtering technique [20]. IP traceback is a challenging problem...
because of the distributed anonymous nature of DDoS attacks, the stateless nature
of the Internet, the destination oriented IP routing, and the fact of having millions of
hosts connected to the Internet (implying huge search space). All these factors help
attackers to stay behind the scenes and hence complicate the process of traceback.
Generally, traceback schemes provide spoofed packets traceback capability either
by augmenting the packets with partial path information (i.e., packet marking), or by
storing packet digests or signatures at intermediate routers (i.e., packet logging).
Such approaches require either a large number of attack packets to be collected by
the victim to infer the paths (packet marking), or a significant amount of
resources to be reserved at intermediate routers (packet logging). This paper adopts
a hybrid traceback approach in which packet marking and packet logging are
integrated to achieve the best of both worlds (i.e., small number of attack packets to
conduct the traceback process, and small amount of resources to be allocated at intermediate
routers (packet logging). Based on this notion, two traceback schemes are presented.

II. RELATED WORK

We refer to a router with high speed links as a high-speed router. We also term a packet
of interest an attack packet. Similarly, the source and destination of an attack packet
is an attacker and a victim, respectively. The sequence of routers traversed by an
attack packet on its way from source to destination make up an attack path. The
attack path from the attacker to the victim is represented as an ordered list of routers
\((R_1; R_2; \ldots; R_m)\). The objective of IP traceback is to figure out this ordered list of
routers. The process of constructing attack paths is called traceback process. Based on
the vulnerability that is exploited, DoS attacks can be classified into flooding attacks
and software exploits \([4]\). Flooding attacks (e.g., smurf, SYN flood) work by flooding
victims with large amounts of traffic. Flooding attacks consume some limited
resource (e.g., link bandwidth or computing resource) at victims to prevent legitimate
users from accessing that resource. Software exploits (e.g., teardrop, ping-of-
death) work by sending victims a single or a few malformed packets to abuse some
feature or implementation bug of operating systems or applications to disable the
service. In our view, this is equivalent to identifying the end point of a link list starting
at the victim, where each element in the list represents an intermediate router along the
path from victim to attacker as can be seen in Fig. 1. Multiple attackers case
corresponds to a tree of link lists rooted at the victim ( ), where each leaf represents a
link list end point.

Fig1. An Instance of the Trace-back

The main assumptions made in our work are similar to those made in \([15, 17] \) and
\([6]\), with an exception that we do not necessarily assume that each attack source
has to send numerous packets. The imminent threats imposed by DoS attacks

call for efficient and fast traceback schemes
that enjoy the following features: 1) Providing accurate information about routers near the attack source rather than those near the victim. 2) Recognition and exclusion of false information injected by the attacker. 3) Avoiding the use of large amount of attack packets to construct the attack path or attack tree. 4) Low processing and storage overhead at intermediate routers. 5) Efficient collection of marking information stored at intermediate routers (if any). Previous schemes failed to satisfy the above features collectively. For example, in PPM [15], routers that are far away from the victim have very low chance to pass their marking information to the victim because intermediate routers overwrite this information, which leads to the loss of valuable marking information written by routers near the attacker(s). This is contradictory to our goal of having more knowledge about these routers. PPM requires considerable amount of packets to be collected at the victim before conducting the traceback process. Waiting for a large number of attack packets to be collected at the victim will significantly increase the response time of counting an attack. This argument may not be valid when a small value for the marking probability is used in PPM. However, such marking probability introduces a serious vulnerability in PPM that was pointed out in [14], where the attacker has the ability to pass spoofed marking information to mislead the victim. The major drawback of Hash-based traceback [19] is that it incurs a heavy burden on routers by requiring them to log information about every forwarded packet. Moreover, the method employed to collect packet information from network routers is inefficient and requires special resources; the scheme assumes that a central management unit is available in each domain to download and search all packet digests looking for specific packet footprints. This results in a tedious process and an unnecessary overhead. The deterministic storage algorithm increases the memory requirement of the scheme. Moreover, a major concern in Hash-based traceback is the small window of time through which packets can be successfully traced. Generally, routers write their IP addresses in the forwarded packets under the sampling based schemes. In contrast, packet information (digests or signatures) is written in router’s memory under the logging based schemes. Each approach has its own pros and cons. We believe that developing a hybrid of both approaches, as predicted in [5], can lead to an improved traceback capability. Different than earlier work, we focus on key aspects of an efficient traceback scheme. In particular, we show how to provide the victim with more information about the path followed by an attack packet using constant space in the packet header, which has the benefit of reducing the number of packets required to localize attack sources. Also, we address the issue of eliminating the attacker’s ability to mislead the victim in the traceback process. Also, we focus on reducing the router’s storage requirement. In this paper, we propose two alternative implementations of a hybrid marking and logging based IP traceback. The first implementation is called Distributed Link-List Traceback (DLLT). DLLT has an efficient implementation that combines the good features of both PPM and Hash-based schemes based on the idea of preserving the marking information at intermediate routers in such a way that it can be collected using a link-list based approach. The second implementation is called Probabilistic Pipelined Packet Marking (PPPM). This scheme aims at propagating the IP addresses of the routers that were involved in marking certain packet by loading them into packets going to the same destination. Therefore, preserving these addresses while avoiding the need for long term storage at intermediate routers. DLLT and PPPM exhibit the features of PMM [15] in the sense that routers mark forwarded packets probabilistically. Also, they exhibit the features of Hashbased scheme [19] in the sense that processing and storage at intermediate routers are
necessary. The main advantage of PPPM scheme over the DLLT scheme is that long term storage at intermediate routers is not necessary.

III. DISTRIBUTED LINK LIST TRACEBACK

The main idea of DLL is to keep track of a subset of the routers that are involved in forwarding certain packet by establishing a temporary link between them in a distributed manner. DLL is based on a "store, mark and forward" approach. A fixed-size marking field is allocated in each packet. Any router that decides to mark the packet, stores the current content of the marking field (which was written by the previous marking router) in a special data structure called Marking Table maintained at the router. The router generates an ID for that packet to index its marking information in the marking table. The router marks the packet by overwriting the marking field by its own IP address, and then forwards the packet as usual. Any router that decides not to mark the packet just forwards it. A link list is inherently established because the marking field serves as a pointer to the last router that did the marking for a given packet, and the marking table of that router contains a pointer (i.e., the IP address) to the previous marking router, and so on. Therefore, each packet received by the destination contains the start point of a link list that is part of the packet path. We call it distributed link-list because each router decides by its own to be on the list or not according to certain marking probability. The basic components of this scheme are the marking and storage procedure, and the marking information collection protocol. What follows is a discussion of each of these components.

1) Marking and Storage Procedure: DLLT employs a probabilistic marking and storage scheme. When a router receives a packet, it will mark the packet with probability, \( _p \). If the packet has been marked previously, then the router will store that information before remarking the packet. Therefore, packet marking and storage is an integrated procedure. Before going into details of this procedure, we show the main data structure used for storing packet information. Logging packet information at intermediate routers is not a new idea. Storing the packet features was considered in [2]. Also storing packet digests was considered in [19]. The major drawback of these schemes, as pointed, is that they put a heavy burden on routers by requiring them to log information about every forwarded packet. Moreover, there is a need to download and search all this information looking for specific packet footprints, which results in a tedious and an unnecessary process. Our storage scheme is probabilistic in nature, which means that only fraction of the traffic is to be logged at each router. Also, we store this information in such a way as to ensure that it can be collected in a predetermined manner.

IV. Security of the Proposed DLLT

DLLT security: The correctness of the DLLT scheme is based on the validity of the information collected by the MIC protocol. It is possible that attackers may abuse this protocol in different ways. The protocol can be abused by generating fake MIC-requests and/or fake MIC-replies. This is inevitable as it is inherent problem with any protocol and it is not specific to the proposed MIC protocol. Fake MIC-requests may increase processing overhead at network routers. However, they do not affect the outcome of the traceback process. Fake MIC-replies, on the other hand, can affect the traceback outcome, because such replies inject false information about the attack path/tree. To overcome this problem, the MIC protocol can be modified in such a way to provide light weight authentication of MIC-replies. A straightforward modification is to let the victim attach a secret code to each generated MIC-request. This code propagates with each MIC-request as it becomes part of the request. The same secret code is copied into the corresponding MIC-reply. Fake MIC-replies can be easily filtered as they are not supposed to contain a valid code. For
increased security, the secret code should be distinct for each MIC-request. The deadend router which may be chosen in advance by the attacker can be compromised. For example: (i) it can create looping by propagating an MIC-request packet to one of the routers found in the Marking Routers List of that packet (ii) it may ignore any MIC-request packet (iii) it may propagate received MIC-request packets to wrong routers. Issue (i) can be avoided by preventing a router from propagating a MIC-request packet if its address is already in the Marking Routers List. Issues (ii) and (iii) can be detected by modifying the MIC protocol such that each router acknowledges the requesting router. If no acknowledgment is received within a specific amount of time. The requesting router detects that something is wrong, identifies itself to be the end-list router, and generates the MIC-reply packet. **PPPM Security:** The PPPM scheme is vulnerable to the following types of attacks. The attacker is expected to behave like a marking router. Therefore, marking its outgoing packets with false information before being marked by any router along their path. For example, it can spoof the IP-marking field. Also, it can write any value in the field. If the attacker uses a value that is less than the original packet’s TTL, then it is easy to distinguish and drop attack packets before reaching the victim, because must be larger than the TTL value of a given packet. If the written by the attacker is larger than the packet’s original TTL, then intermediate routers cannot distinguish attack packets. This false information is more likely to be overwritten by subsequent routers if a reasonable marking probability is used. Unfortunately, there is nothing that can prevent the false information from propagating to the victim. For a given packet ID, the victim will include the spoofed marking router in the marking routers list for that ID. However, because of the restriction that we have imposed on be larger than TTL, the distance obtained at the victim for the spoofed marking router will always be larger than that obtained for any other marking router for the same packet. Therefore, the victim can easily distinguish and exclude this information. As another vulnerability of the PPPM scheme, the attacker may use the same packet ID for all outgoing packets. Worse than that, multiple attackers can coordinate with each other to use the same packet ID either continuously or from time to time. In PPPM, the first marking router cannot recognize itself as the first router to do the marking for a given packet. Therefore, the ID injected by the attacker is used. We suggest the following slight modification to the marking and buffering to prohibit attackers from using the same ID for all packets: if a router receives a packet, P, that has identical (P.dest, P.ID) to those buffered for previously marked packet, it will drop the incoming packet. The suggested modification is based on the fact that it is very unlikely for two distinct packets traveling to the same destination within small window of time to have identical IDs.

**V. Performance Evolution**

The effectiveness of log-based IP traceback increases greatly with the widespread deployment of traceback-enabled routers in the network. Similar to SPIE, it is likely that hybrid single-packet IP traceback does not require all routers to be traceback-enabled. All traceback-enabled routers within a network can be regarded as an overlay network. If the traceback server has the topology knowledge of that overlay network and each traceback-enabled router knows its overlay neighbors, the hybrid approach still works. We evaluate the performance of the proposed schemes through a combination of theoretical and simulation studies. Theoretically, we bound the number of packets required by DLLT/PPPM for full path construction and compare it to that required by the PPM scheme [15], and we quantify the storage requirement of DLLT and PPPM. We evaluate the performance of the proposed schemes through extensive simulation experiments against various performance metrics. It is obvious that the proposed schemes are similar in
functionality, in the sense that marking information that belongs to a certain packet is preserved to be later collected by the victim. While the number of packets required to conduct a full attack path is different in both schemes (few additional packets are required by the PPPM scheme to flush the pipeline), the number of packets that must be sent by an attacker such that every router along the path to the victim is involved in marking at least one of these packets is the same in both schemes (assuming the same _ in both cases). Our objective is to find a bound on the minimum number of packets that has to be sent by the attacker such that every router along the path to the victim is involved in marking at least one of these packets with high confidence probability . The probability that a packet will be marked by a router and then left unmarked by all downstream routers is a strictly decreasing function of the distance to the victim in the case of PPM. Therefore, increasing the path length reduces the chance to obtain marks from far away routers. Hence, more packets are required to ensure that all routers along the path are represented by at least one mark. Coming to DLLT/PPPM, the number of packets required for traceback does not remain constant as path length increases. In fact, there is a slight increase which is difficult to notice in the plot because it is too small compared to the scale of the Y axis. This increase, though slight, is intuitive because more packets would be required to cover a longer path.

VI. Conclusion

Tracing a single IP packet back to its origin is the ultimate goal of IP traceback. SPIE illustrates the feasibility of tracing individual packets with packet logging. An efficient traceback scheme is necessary to identify the sources of DoS attacks which impose an imminent threat to the availability of Internet services. The work presented in this paper adopts a hybrid traceback approach in which packet marking and packet logging are integrated to achieve the best of both worlds (i.e., small number of attack packets to conduct the traceback process, and small amount of resources to be allocated at intermediate routers for packet logging purposes). Based on this notion, two traceback schemes were proposed. The first scheme, called Distributed Link-List Traceback (DLLT), is based on the idea of preserving the marking information at intermediate routers in such a way that it can be collected in an efficient manner. The second scheme, called Probabilistic Pipelined Packet Marking (PPPM), employs the concept of “pipeline” for propagating marking information from one marking router to another so that it eventually reaches the destination. The proposed schemes enjoy many of the features mentioned. Their probabilistic nature of marking and storage offers the advantage of minimizing router's processing and storage overhead. Also, both schemes eliminate attacker's ability to mislead the victim. This is achieved in DLLT by storing the packet digests at intermediate routers, which provides an authentic way to verify that a given router has actually forwarded certain packet. In PPPM, spoofed marking information written by the attacker can be discarded by observing that the distance associated with it is always the largest among distances obtained for marking information that correspond to the same packet. Marking information is collected from intermediate routers efficiently. For example, DLLT collects relevant marking information from specific routers in a predetermined manner using the link list approach. PPPM collects the marking information by loading them into packets going to the same destination.

References


management,” in SIGCOMM, (Cambridge, MA, USA), August 1999.


Mr. V S R S Phani Kumar is a student of Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP, India. Presently he is Pursuing his M.Tech (CN & IS) from this college. His area of interests includes Information Security, Cloud Computing, Data Communication & Networks.

Mr. P. Karunakar Reddy is an efficient teacher, received M.Tech from JNTU Hyderabad is working as an Associate Professor in Department of C.S.E, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam, AP, India. He has published many papers in both National & International Journals. His area of interest includes Data Communications & Networks, Database Management Systems, Computer Organization, C Programming and other advances in Computer Applications.

Mr. S. Suresh well known Author and excellent teacher. He is working as a H.O.D for the Department of M.Tech., Computer Science & Engineering, Sree Rama Institute of Technology & Science, Kuppenakuntla, Penuballi, Khammam. He has vast teaching experience in various engineering colleges. To his credit couple of publications both National & International conferences / journals. His area of Interest includes Data Warehouse and Data Mining, information security, Data Communications & Networks and other advances in Computer Applications. He has guided many projects for Engineering Students.