NORTHWESTERN
UNIVERSITY

Electrical Engineering and Computer Science Department

Technical Report
NWU-EECS-07-02
February 3, 2006

Network-based and Attack-resilient Length Signature Generation for Zero-day Polymorphic Worms

Zhichun Li, Lanjia Wang, Yan Chen and Zhi (Judy) Fu

Abstract

It is crucial to detect zero-day polymorphic worms and to generate signatures at the edge network gateways or honeynets so that we can prevent the worms from propagating at their early phase. However, most existing network-based signatures generated are not vulnerability based and can be easily evaded under attacks. In this paper, we propose to design vulnerability based signatures without any host-level analysis of worm execution or vulnerable programs. As the first step, we design a network-based Length-based Signature Generator (LESG) for worms based on buffer overflow vulnerabilities. The signatures generated are intrinsic to buffer overflows, and are very hard for attackers to evade. We further prove the attack resilience bounds even under worst case attacks with deliberate noise injection. Moreover, LESG is fast, noise-tolerant, and has efficient signature matching. Evaluation based on real-world vulnerabilities of various protocols and real network traffic demonstrates that LESG is promising in achieving these goals.

Keywords: zero-day vulnerability; polymorphic worm; worm signature generation; network intrusion detection/prevention system (IDS/IPS); protocol field length based signature

Network-based and Attack-resilient Length Signature Generation for Zero-day Polymorphic Worms

Zhichun Li, Lanjia Wang†, Yan Chen and Zhi (Judy) Fu‡
Northwestern University, Evanston, IL, USA
†Tsinghua University, Beijing, China
‡Motorola Labs, Schaumburg IL, USA

Abstract

It is crucial to detect zero-day polymorphic worms and to generate signatures at the edge network gateways or honeynets so that we can prevent the worms from propagating at its early phase. However, most existing network-based signatures generated are not vulnerability based and can be easily evaded under attacks. In this paper, we propose to design vulnerability based signatures without any host-level analysis of worm execution or vulnerable programs. As the first step, we design a network-based Length-based Signature Generator (LESG) for worms based on buffer overflow vulnerabilities[1]. The signatures generated are intrinsic to buffer overflows, and are very hard for attackers to evade. We further prove the attack resilience bounds even under worst case attacks with deliberate noise injection. Moreover, LESG is fast, noise-tolerant, and has efficient signature matching. Evaluation based on real-world vulnerabilities of various protocols and real network traffic demonstrates that LESG is promising in achieving these goals.

1 Introduction

Attacks are commonplace in today's networks, and identifying them rapidly and accurately is critical for large network/service operators. It was estimated that malicious code (viruses, worms and Trojan horses) caused over $28 billion in economic losses in 2003, and will grow to over $75 billion in economic losses by 2007 [2]. The intrusion detection systems (IDSes) [3, 4] are proposed to defend against malicious activities by searching the network traffic for known patterns, or *signatures*. So far such signatures for the IDSes are usually generated manually or semi-manually, a process too slow for defending against self-propagating malicious codes, or *worms*.

Thus, it is critical to automate the process of worm detection, signature generation and signature dispersion in the early phase of worm propagation, especially at the network level (gateways and routers). There is some existing work towards this direction [5–7].

However, to evade detection by signatures generated with these schemes, attackers can employ *polymorphic* worms which change their byte sequence at every successive infection. Recently, some polymorphic worm signature generation schemes are proposed. Based on characteristics of the generated signatures, they can be broadly classified into two categories – *vulnerability-based* and *exploit-based*. The former signature is inherent to the vulnerability that the worm tries to exploit. Thus it is independent of the worm implementation, unique and hard to evade, while exploit-based signatures captures certain characteristics of a specific worm implementation. However, schemes of both categories have their limitations.

Existing vulnerability-based signature generation schemes are host-based and cannot be applied for detection at the network router/gateway level. These schemes [8–10] either require exploit code execution or the source/binary code of the vulnerable program for analysis. However, such host-level schemes are too slow to counteract the worms that can propagate at exponential speed. Given rapid growth of network bandwidth, today's fast propagation of viruses/worms can infect most of the vulnerable machines on the Internet within ten minutes [11] or even less than 30 seconds with some highly virulent techniques [12, 13] at near-exponential propagation speed. At the early stage of worm propagation, only a very limited number of

[1] It is reported that more than 75% of vulnerabilities are based on buffer overflow [1].

worm samples are active on the Internet and the number of machines compromised is also limited. Therefore, signature generation systems should be network-based and deployed at high-speed border routers or gateways where the majority of traffic can be observed. Such requirement of network-based deployment severely limits the design space for detection and signature generation as discussed in Section 2.

Existing exploit-based schemes are less accurate and can be evaded. Some of these schemes are network-based and are much faster than those in the former category. However, most of the schemes are content-based which aim to exploit the residual similarity in the byte sequences of different instances of polymorphic worms [14–18]. However, as mentioned in [18], there can be some worms which do not have any content-based signature at all. Furthermore, various attacks have been proposed to evade the content-based signatures [19–22]. The rest of them [23, 24] generate signatures based on exploit code structure analysis, which is not inherent to the vulnerability exploited and can also be evaded [19].

Therefore, our goal is to design a signature generation system which has both the accuracy of vulnerability-based scheme and the speed of exploit-based scheme so that we can deploy it at the network level to thwart zero-day polymorphic worm attacks. As the first step towards this ambitious goal, we propose LEngth-based Signature Generator (called $LESG$) which is a network-based approach for generating efficient and unevadable length-based signatures. That is, even when the attacker know what the signatures are and how the signatures are generated, they still cannot find efficient and effective way to evade the signatures.

Length-based signatures target buffer overflow attacks which constitutes the majority of attacks [1]. The key idea is that in order to exploit any buffer overflow vulnerabilities, the length of certain protocol fields must be long enough to overflow the buffer. A buffer overflow vulnerability happens when there is a vulnerable buffer in the server implementation and some part of the protocol messages can be mapped to the vulnerable buffer. When an attacker injects an overrun string input for the particular field of the protocol to trigger the buffer overflow, the length of such input for that field is usually much longer than those of the normal requests. Thus we can use the field input length to detect attacks. This is intrinsic to the buffer overflow, and thus it is very hard for worm authors to evade.

In addition to being network based and having high accuracy, LESG has the following important features.

- **Noise tolerance.** Signature generation systems typically need a flow classifier to separate potential worm traffic from normal traffic. However, network-level flow classification techniques [7, 25–28] invariably suffer from false positives that lead to noise in the worm traffic pool. Noise is also an issue for honeynet sensors [5, 16, 23]. For example, attackers may send some legitimate traffic to a honeynet sensor to pollute the worm traffic pool and to evade noise-intolerant signature generation. Our LESG is proved to be noise tolerant or even stronger, attack resilient, i.e. LESG works well with maliciously injected noise in an attempt to mislead NIDS [19].
- **Efficient Signature Matching.** Since the signatures generated are to be matched against *every flow* encountered by the NIDS/firewall, it is critical to have fast signature matching algorithms. Moreover, for the network-level signature matching, the signatures must be based solely on the network flows rather than host-level information such as system calls. In LESG system, with a protocol length parser, the length-based signature can be matched at network level without any host-level analysis. That is, we can directly check the packets against signatures at routers/gateways.

In the rest of the paper, we first survey related work in Section 2 and discuss the LESG architecture in Section 3. Then we present the length-based signature generation problem in Section 4, generation algorithm in Section 5, and its attack resilience in Section 6. After that, in Section 7, we use real Internet traffic and seven real exploit code (enhanced with polymorphic capabilities) on five different protocols to test the performance of LESG prototype. Results show that LESG is highly accurate, noise tolerant, capable of detecting multiple worms in the same protocol pool, and capable of online signature generation with small memory consumption. Finally, we discuss some practical issues in Section 8 and conclude in Section 9.

2 Related Work

Early automated worm signature generation efforts include Honeycomb [5], Autograph [7], and EarlyBird [6]. But they do not work well with polymorphic worms.

Property of signatures generated	Signature generation mechanisms	
	Network-based	Host-based
Exploit-based	Polygraph [15], Hamsa [14], PADS [16], Nemean [23], CFG [24]	DACODA [18], Taint check [17]
Vulnerability-based	**LESG**	Vulnerability signature [10], Vigilante [29], COVERS [8], Packet Vaccine [9]

Table 1: Comparison with other polymorphic worm signature generation schemes.

As mentioned before, existing work on automated polymorphic worm signature generation can be broadly classified into vulnerability-based and exploit-based. Based on signature generation input requirements, we can further categorize these schemes on another axis: host-based vs. network-based. The former requires either exploit code execution or the source/binary code of the vulnerable program for analysis. On the other hand, the network based approach relies solely on network-level packets. The classification of existing schemes and LESG is shown in Table 1. We discuss them in more details below.

Exploit-based schemes. We have discussed most of them in the introduction [14–18, 23, 24]. For example, Christopher *et al.*proposes using structural similarity of Control Flow Graph (CFG) to generate a fingerprint to detect different polymorphic worms [24]. However, their approach can be evaded when the worm body is encrypted. Furthermore, compared with length-based signatures, it is much more computationally expensive to match the fingerprint with the network packets. Thus it cannot be applied to filter worm traffic on high-speed links.

Compared with some most recent work in this category, such as Hamsa [14], LESG has better attack resilience. For example, it has better bounds for the deliberate noise injection attacks [19].

Vulnerability-based and host-based schemes. Brumley *et al.*presents the concept of vulnerability signature in [10] and argues that the best vulnerability signatures are Turing machine signatures. However, since the signature matching for Turing machine signatures is undecidable in general, they reduce the signatures to symbolic constraint signatures or regular expression signatures. Their approach is a heavy-weight host-based approach, which has high computation overhead and also needs some information such as the vulnerable program, multiple execution traces, and the vulnerability condition. Similarly , Vigilante [29] proposed a vulnerability based signature which is similar to the MEP symbolic constraint signatures in [10].

Liang *et al.*proposed the first host-based scheme to generate length-based signatures [8]. Packet Vaccine [9] further improve the signature quality by using binary search. Unfortunately, both of them are host-based approaches and are subject to the limitations mentioned before and some additional shortcomings. First, they need to know the vulnerable program. Sometimes, they have to try many different implementation versions to find the vulnerable ones. Second, the signature generated by [8] based on a small number of samples may be too specific to represent the overall worm population. Therefore, detection based on their generated signatures tends to have high false negatives. Moreover, the protocol specification language used in their approach is not expressive enough for many protocols.

Other related work. There are previous research efforts on network-level detection of buffer overflow exploits. However, they do not generate any effective signatures for checking future traffic for worms due to high matching overhead and high false positives. Buttercup [30] and TCTP [31] detect buffer overflow attacks by recognizing jump targets within the sessions. Approaches like SigFree [32] and [33] detect exploit codes based on control flow and data flow analysis.

3 Architecture of LESG

As shown in Figure 1, $LESG$ can be connected to multiple networking devices, such as routers, switches and gateways. Most modern switches are equipped with a span port to which copies of the traffic from a list of ports can be directed. $LESG$ can use such a span port for monitoring all the traffic flows. Alternatively, we can use a splitter such as a Critical Tap [34] to connect $LESG$ to routers. Such splitters are fully passive and used in various NIDS systems to avoid affecting the traffic flows. In addition, $LESG$ can also be used to monitor traffic towards large-scale honeynet/honeyfarm through sniffing the traffic on its gateways. Nowadays, there are some large honeynets even with /A network size [35–37].

Similar to the basic framework of Polygraph [15] and Hamsa [14], we first need to sniff the traffic from networks, and classify the traffic to different application level protocols based on port numbers or other protocol identifiers. Then we can filter out known worms and then further separate the traffic to form a suspicious traffic pool and a normal traffic reservoir using an existing flow classifier [7, 25–28]. The flow classifier is also similar to the one in Polygraph [15] and Hamsa [14] system, and can integrate various techniques (such as honeypot/heneynet, port scan detection, and other advanced techniques) to identify suspicious flows. Note that the flow classifiers can also operate with line speed of routers as achieved in our earlier work with scan detection [38]. It is called suspicious pool rather than malicious pool because the behavior based classification can never be perfectly accurate.

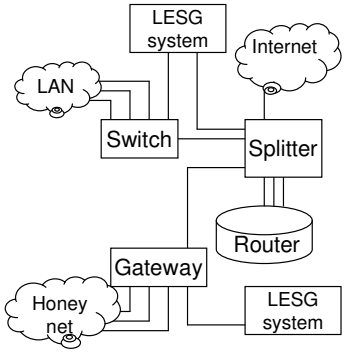

Figure 1: Deployment of LESG.

Leveraging the normal traffic selection policy mentioned in [14], we can create the normal pool. The suspicious pool and the normal pool are inputted to the signature generator as shown in Figure 2. We first specify the protocol semantics and use a protocol parser to parse each protocol session into a set of fields. Each field is associated with a length and a type. The field length information of both the suspicious pool and the normal pool are given as input to the "LESG core"(signature generation algorithm) module to generate the signatures.

3.1 Protocol Parsing

As emphasized in [39], protocol parsing is an important step to any semantic analysis of network traffic, such as network monitoring [40], network intrusion detection system [3, 4], smart firewalls, *etc.*. We analyzed three text-based protocols (HTTP, FTP, and SMTP) and 7 binary protocols (DNS, SNMP, SMB, WINRPC, SUNRPC, NTP, SSL). We find, in general, it is much easier to parse the lengths of the protocol fields than full protocol parsing.

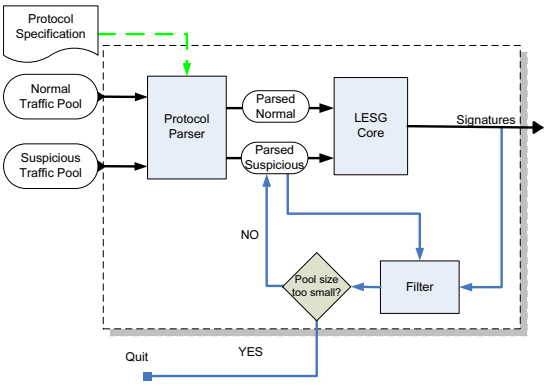

Figure 2: LESG signature generator

Some recent researches, such as BINPAC [39], have studied how to ease the job of writing a protocol parser. BINPAC is a yacc like tool for writing application protocol parsers. It has a declarative language and compiler, and actually works as a parser generator. Its input is a script which is actually a protocol specification written in BINPAC language. The output is a parser code for that protocol. Currently BINPAC is executed in connection with Bro [4] which implements other necessary traffic analysis at lower levels. With BINPAC, writing a protocol parser has been greatly simplified. Furthermore, not only the available scripts provided by Bro can be reused, but also many people can potentially contribute to produce more reusable protocol specifications for BINPAC as an open source tool. Because of these advantages, we use BINPAC and Bro for packet flow reassembling and protocol parsing in our research.

4 Length-Based Signature Definition and Problem Statement

In this section, we formally model each application message as a field hierarchy, and present it as a vector of fields. Based on this model, we formally define the length-based signatures and the length-based signature generation problem.

4.1 Field Hierarchies

Each of the application sessions (flows) usually contains one or more Protocol Data Units (PDUs), which are the atomic processing data units that the application sends from one endpoint to the other endpoint. PDUs are normally specified in the protocol standards/specifications, such as RFCs. A PDU is a sequence of bytes, and can be dissected into multiple *fields*. Here, a field means a sub-sequence of bytes specified in the protocol

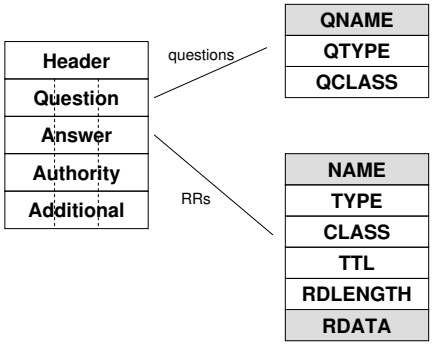

Figure 3: Illustration of DNS PDU

Figure 4: Abstraction of DNS PDU

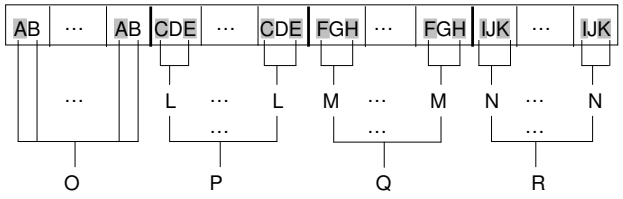

Figure 5: Hierarchical Structure of DNS PDU

standard, having certain meaning or functionality for the protocol. Typically, a field encodes a variable with a certain data structure, such as a string, an array *etc.*. Take the DNS protocol as an example, figure 3 shows the format of the DNS PDUs [41]. It has a header and four other sections – QUESTION, ANSWER, AUTHORITY and ADDITIONAL. Each section is further composed of a set of fields. The QUESTION section contains one or more DNS queries that are further composed of field class QNAME, QTYPE and QCLASS. Another three sections contain one or more Resource Records (RRs), and each RR is composed of six lower level fields (NAME, TYPE, *etc.*). Borrowing similar terms from the object model, we call the type of fields, such as QNAME and QTYPE, as the *field class*, and each concrete instance of certain field as an *instance* of the field.

Among all the field classes in PDUs, some, *e.g.*, QNAME, NAME and RDATA, are *variable-length fields*, whose instances possibly have different lengths; others are *fixed-length fields*, whose instances all have the same length which is defined in the protocol standard. In our analysis, the continuous fixed-length fields can be combined as one field for simplicity. Again, using the DNS protocol as an example, we denote the variable-length field QNAME as A, and the concatenation of field QTYPE and QCLASS as B because both of them are fixed-length fields. Then we denote the variable-length field NAME in section ANSWER as C, and the concatenation of the next four fields as D because they are all fixed-length fields. With these abstractions, the DNS PDU is illustrated as Figure 4, where there are totally 11 classes of fields. The 7 fields with gray background are variable-length fields.

We make the following two observations on such a representation of PDU. First, the number of instances of one field class in a PDU may vary. For example, one PDU may contain one instance of field A, and another PDU may contain two. Secondly, in certain server implementations, it is possible that the concatenation of multiple field instances (of the same field class or not) are stored in one buffer. That is, if the server has an overflow vulnerability related to this buffer, it is the concatenation of several field instances that can overflow the buffer. For example, imagine a DNS server receives a DNS PDU and stores the entire PDU in a vulnerable buffer, what overflows the buffer is the concatenation of all the field instances. These two observations have been further validated on other protocols such as SNMP and WINRPC.

With these considerations, we design a hierarchical model to describe the possible field classes in a PDU. As Figure 5 shows, we denote the QUESTION section as a new field O, a concatenation of all the instances of field A and B, $O = (AB)^*$. We also denote the concatenation of field C, D and E as a new field $L = CDE$, the concatenation of all fields L (namely section ANSWER) as another new field $P = L^*$, and so on. In short, we include all possible variable-length fields that potentially correspond to vulnerable buffers. We build such a hierarchy for every flow.

In the rest of the paper, for brevity, we refer to variable-length fields simply as fields. Suppose there are totally K classes of fields in the hierarchy constructed for a certain protocol. We use an index set $E = \{1, 2, \ldots, K\}$ to denote these K fields. Let $x_k, k = 1, 2, \ldots, K$, be the maximum among the lengths of potentially multiple instances of field k, then a vector $X = (x_1, x_2, \ldots, x_K)$ is generated to represent the field lengths for each field in a session (flow).

$\mathcal{M}$: suspicious traffic pool	$\mathcal{N}$: normal traffic pool
$\|\mathcal{M}\|$: number of suspicious flows in $\mathcal{M}$	$\|\mathcal{N}\|$: number of noise flows in $\mathcal{N}$
$\mathcal{M}^1$: set of true worm flows in $\mathcal{M}$	$\mathcal{M}^2$: set of noise flows in $\mathcal{M}$
α : coverage of true worms	K : number of variable length fields
$\mathcal{M}_\mathcal{S}$: set of suspicious flows covered by signature set $\mathcal{S}$	$\mathcal{N}_\mathcal{S}$: set of normal flows covered by signature set $\mathcal{S}$
$\mathrm{COV}_\mathcal{S}$: $\frac{\|\mathcal{M}_\mathcal{S}\|}{\|\mathcal{M}\|}$ for a signature set $\mathcal{S}$	$\mathrm{FP}_\mathcal{S}$: $\frac{\|\mathcal{N}_\mathcal{S}\|}{\|\mathcal{N}\|}$ for a signature set $\mathcal{S}$
COV_0 : minimum coverage requirement for a signature candidate	FP_0 : maximum false positive ratio for a signature candidate
γ' : minimum coverage increase requirement for a signature to be outputted in the first loop of the Step 3 algorithm	γ : minimum coverage increase requirement for a signature to be outputted in the second loop of the Step 3 algorithm

Table 2: Table of Notations

4.2 Length-based Signature Definition

Based on the length vector representation of a session, we formally define the concept of *length based signature*. A signature is a pair $S_j = (f_j, l_j)$, where $f_j \in E$, f_j is the signature field ID, and l_j is the corresponding signature length for field f_j. When using the signature to detect the worms, the matching process is as follows. For a flow $X = (x_1, x_2, \ldots, x_K)$, we compare x_{f_j} and l_j. If $x_{f_j} > l_j$, then the flow X is labelled as a worm flow; otherwise it is labelled as a normal one.

More than one signature correspond to different fields can possibly be generated for a given protocol, resulting in a *signature set* $S = \{S_1, S_2, \ldots, S_J\}$. A flow, which may contain one or more PDUs, will be labelled as worm if it is matched by at least one signature in the set.

The length based signatures are designed for buffer overflow worms. The signature field should be exactly mapped to a vulnerable buffer. In this case, the field of this instance must be longer than the buffer to overflow it, while normal instances must be shorter than the buffer. Note that different servers may implement different buffer lengths if the maximal length is not specified in the RFC. Here we focus on popular implementations because the spread speed and scope of worms will be significantly limited if they only target unpopular implementations. We define the minimum buffer length of popular implementations as *the ground truth signature*, denoted as $B = (f_B, L_B)$ where L_B is the vulnerable buffer length. Even with multiple different implementations, for the field related to the vulnerable buffer, the distributions of normal flows and worm flows should be well apart. That is, the lengths of normal flows should be less than L_B because for a popular server implementation (*e.g.*, FTP), there are often various client softwares communicating with it without knowing its buffer length. So L_B should be large enough for most of the normal flows. On the other hand, obviously those of worm flows should be larger than L_B.

As elaborated below, our algorithm will not output any signatures for non-buffer-overflow worms because our algorithm ensures that all generated signatures have low false positives.

4.3 Length-Based Signature Generation Problem Formulation

A worm flow classifier labels a flow as either worm or normal. The flows labelled as worms constitute the *suspicious* traffic pool while those labelled normal constitute the *normal* traffic pool. If the flow classifier is perfect, all the flows in the suspicious pool are worm samples. If the worm is a buffer overflow worm, finding a length-based signature amounts to simply finding the best field and the field length with the minimal false negatives and the minimal false positives. However, in practice flow classifiers at the network level are not perfect and always have some false positives and therefore the suspicious pool may have some normal flows. Finding signatures from a noisy suspicious pool makes the problem NP-Hard (Theorem 1). On the other hand, due to the large volume traffic on the Internet, we assume the noise (worm flows) in the normal pool is either zero or very limited, and thus it is negligible.

After filtering existing known worms, there can be multiple worms of a given protocol in the suspicious pool, though the most common case is a single worm having its outbreak undergoing in the newly generated suspicious pool. The output of the signature generation is a signature set $S = \{S_1, S_2, \ldots, S_J\}$. A flow matched by any signature in this set will be labelled as a worm flow.

In Table 2, we define most of the notations used in the problem formulation, theorems, and their proofs.

Problem 1 (Noisy Length-Based Signature Generation (NLBSG))**.**
INPUT*: Suspicious traffic pool* $\mathcal{M} = \{M_1, M_2, \ldots\}$ *and normal traffic pool* $\mathcal{N} = \{N_1, N_2, \ldots\}$*; value* $\gamma < 1$.
OUTPUT*: A set of length-based signature* $\mathcal{S} = \{(f_1, l_1), \ldots, (f_J, l_J)\}$ *such that* $\mathrm{FP}_{\mathcal{S}}$ *is minimized subject to* $\mathrm{COV}_{\mathcal{S}} \geq 1 - \gamma$.

Hardness For a buffer overflow worm in the suspicious pool, in the absence of noise, generation of the set of length-based signatures is a polynomial time problem, since we know the size of the set is one. However, with noise and multiple worms, the computational complexity of the problem has significantly changed.

Theorem 1. *NLBSG is NP-Hard*

Proof Sketch. The proof is by reduction from Minimum k Union, which is equivalent to Maximum k-Intersection [42]. □

5 Signature Generation Algorithm

Although, the problem NLBSG is NP-Hard in general, for buffer overflow worms, the algorithms we proposed are fast and can have fair accuracy even in the worst case scenarios. We formally proved the theoretical false positive and false negative bounds with or without adversaries to inject intentionally crafted noise. To the best of our knowledge, we are the *first* network based signature generation approach that has the accuracy bound even with adversaries' injected noise.

The protocol parsing step generates (field id. length) pairs for all flows in normal traffic pool and suspicious traffic pool respectively. Based on that, we design a three-step algorithm to generate length-based signatures.

Step 1: ***Field Filtering*** Select possible signature field candidates.

Step 2: ***Signature Length Optimization*** Optimize the signature lengths for each field.

Step 3: ***Signature Pruning*** find the optimal subset of candidate signatures with low false positives and false negatives.

5.1 Field Filtering

In this step of the algorithm, we make the first selection on the fields that are possible to be signature candidates. The goal is to limit the searching space. Two parameters are set as the input: FP_0 and COV_0, which indicates the most basic requirement on the false positives and detection coverage. For example, in our experiments, we choose $\mathrm{FP}_0 = 0.1\%$ and $\mathrm{COV}_0 = 1\%$.

In the algorithm below, $\mathcal{N}_{l_j}$ and $\mathcal{M}_{l_j}$ denote the flows detected by signature (f_j, l_j) in pool $\mathcal{N}$ and $\mathcal{M}$ respectively.

Given a parsed normal pool and suspicious pool with sets of (field id. length) pairs, we first sort lengths for every field for both normal pool and suspicious pool respectively. In this first step of algorithm, initially the candidate signature set is empty. Then in the loop, from normal pool, for each field f_j, we find a length so that the normal flows falsely detected by the length signature is less than or equal to FP_0 and the normal flows detected by a shorter length signature will be greater than FP_0. Then in the second part within the loop, we check if that length has detection coverage greater than the minimal coverage COV_0. Therefore, in this first step, the algorithm added all possible candidates that meet the most basic requirements of FP_0 and COV_0. We set low FP_0 as basic low-false-positive requirement. A conservatively small value is chosen for COV_0 initially because attackers may inject a lot of noise into suspicious pool. We

Algorithm **Step 1** *Field filtering* ($\mathcal{M}$,$\mathcal{N}$)

$S \leftarrow \emptyset$;
for field $f_j = 1$ to K
  find l_j such that $\frac{|\mathcal{N}_{l_j}|}{|\mathcal{N}|} \leq \mathrm{FP}_0 < \frac{|\mathcal{N}_{l_j-1}|}{|\mathcal{N}|}$;
  if $\frac{|\mathcal{M}_{l_j}|}{|\mathcal{M}|} \geq \mathrm{COV}_0$
    $S \leftarrow S \cup \{(f_j, l_j)\}$;
  end
end
Output S ;

will further optimize the values in the subsequent steps.

We process each field class separately. According to FP_0, an signature length can be determined, by an sorting and searching, in $O(|\mathcal{N}| \log |\mathcal{N}|)$ time. If the corresponding detection coverage on $\mathcal{M}$ is larger than COV_0, this field is taken as a signature candidate, and is passed to the next step of algorithm, which can be determined by $O(|\mathcal{M}|)$. The running time is $O(K|\mathcal{N}| \log |\mathcal{N}| + K|\mathcal{M}|)$. Since usually $|\mathcal{M}|$ is far smaller than $|\mathcal{N}|$, the overall time cost is $O(K|\mathcal{N}| \log |\mathcal{N}|)$.

This step actually makes use of the fact that, for buffer overflow worms, the true worm samples should have longer lengths on the vulnerable fields than the normal flows, and the noise in suspicious pool that is not injected by attackers should have similar length distribution to traffic in $\mathcal{N}$. If the coverage α of true worm samples in the suspicious pool $\mathcal{M}$ is more than COV_0, the vulnerable field length with small false positive ratio FP_0, should have coverage larger than COV_0 in the suspicious pool. The COV_0 and FP_0 are the very conservative estimate of the coverage and the false positive of the worm.

5.2 Signature Length Optimization

Algorithm **Step 2** *Signature Length Optimization* (S,$\mathcal{M}$,$\mathcal{N}$,$Score(\cdot,\cdot)$)

for signature $(f_j, l_j) \in S$
  sort $\mathcal{M}^{f_j}$ in ascending order;
  find m_0 such that $x^{f_j}_{m_0-1} < l_j < x^{f_j}_{m_0}$;
  $max\_score \leftarrow 0$;
  for $m' = m_0$ to $|\mathcal{M}|$
    $l'_j \leftarrow x^{f_j}_{m'} - 1$;
    if $(max\_score < Score(\mathrm{COV}_{l'_j}, \mathrm{FP}_{l'_j}))$
      $max\_score \leftarrow Score(\mathrm{COV}_{l'_j}, \mathrm{FP}_{l'_j})$;
      $l_j \leftarrow l'_j$;
      $m \leftarrow m'$;
    end
  end
  while $(l_j > \frac{x^{f_j}_{m-1}+x^{f_j}_{m}}{2})$
    if $(Score(\mathrm{COV}_{l_j}, \mathrm{FP}_{l_j}) == Score(\mathrm{COV}_{l_j-1}, \mathrm{FP}_{l_j-1}))$
      $l_j \leftarrow l_j - 1$;
    else
      update S with l_j; break;
    end
  end
end
Output S ;

Algorithm **Step 3** *Signature Pruning* (S,$\mathcal{M}$,$\mathcal{N}$)

$m \leftarrow |\mathcal{M}|$; $\Omega \leftarrow \emptyset$;
$S_1 \leftarrow \{e | e \in S; \mathrm{FP}_e = 0\}$; $S_2 \leftarrow \{e | e \in S; \mathrm{FP}_e > 0\}$;
LOOP1:
while $(S_1 \neq \emptyset)$
  Find $s \in S_1$ such that $\frac{|\mathcal{M}_s|}{m}$ is the maximum one in S_1 ;
  If $(\frac{|\mathcal{M}_s|}{m} \geq \gamma')$
    $\Omega \leftarrow \Omega \cup \{s\}$;$S_1 \leftarrow S_1 - \{s\}$;
    Remove all the samples which match s in $\mathcal{M}$;
  else
    Break ;
  end
end
LOOP2:
while $(S_2 \neq \emptyset)$
  Find $s \in S_2$ such that $\frac{|\mathcal{M}_s|}{m}$ is the maximum one in S_2 ;
  If $(\frac{|\mathcal{M}_s|}{m} \geq \gamma)$
    $\Omega \leftarrow \Omega \cup \{s\}$; $S_2 \leftarrow S_2 - \{s\}$;
    Remove all the samples which match s in $\mathcal{M}$;
  else
    Break ;
  end
end
Output Ω ;

The first step selected candidate signatures to meet the most basic requirements. In the second step, we try to optimize the length value of each candidate signature to improve coverage and to reduce false positives. If the length signature is selected to be too big, there will be less coverage of malicious worm flows. On the other hand, if the length is selected to be too small, there will be a lot of false positives. The first step is a very conservative estimate of coverage. In the second step, we try to find longer length than the first step to improve on coverage without sacrificing false positives. Sometimes a length does improve a lot on coverage of suspicious pool but also increase false positives. We need to have a method to compare different lengths to determine which one is a "better" signature. For brevity, let FP_{l_j} denote the false positive of signature (f_j, l_j), and COV_{l_j} denotes its coverage on $\mathcal{M}$. This step aims to maximize $Score(\mathrm{COV}_{l_j}, \mathrm{FP}_{l_j})$ for each field f_j. We used the notion score function, which is proposed in [14], to determine the best tradeoff between the false positive and coverage. For example, we need to make a choice between $\mathrm{COV} = 70\%$, $\mathrm{FP} = 0.9\%$ and $\mathrm{COV} = 68\%$, $\mathrm{FP} = 0.2\%$.

In the Step 2 algorithm, $\mathcal{M} = \{X_1, X_2, \ldots, X_{|\mathcal{M}|}\}$, where $X_m = (x^1_m, x^2_m, \ldots, x^K_m)$, $m = 1, 2, \ldots, |\mathcal{M}|$ that is the length of each field in a flow m. We define $\mathcal{M}^k = \{x^k_1, x^k_2, \ldots, x^k_{|\mathcal{M}|}\}$. Signature set

S generated in Step 1 is the input of this step.

With the sorted lengths as input, for candidate signature fields, each length above the candiate length selected at step 1 will be tested for its goodness according to the score function, and the best one with the maximum score will be selected. The first loop picks a longer length value with the best score. Then in the second loop, we further optimize it by finding a smaller length with the same score. In $\mathcal{M}^{f_j} = \{x_1^{f_j}, x_2^{f_j}, \ldots, x_{|\mathcal{M}|}^{f_j}\}$, if $x_m^{f_j}$ is in the ascending order, it is easy to know that between any two consecutive elements, namely $x_{m-1}^{f_j}$ and $x_m^{f_j}$, the score is monotonically non-decreasing in l_j. Thus we only need to search among all the $x_m^{f_j} - 1,\ m = m_0, \ldots, |\mathcal{M}|$ for the best score, *i.e.*the total number we need to try is at most $|\mathcal{M}|$.

The rational for the second loop is as follow. Letting the signature length too close to the edge of lengths of worm flows is not a good choice, especially when the length distributions of normal field instances and of malicious field instances are well separated. So in the other part of the algorithm Step 2, l_j decreases until the score changes (decreases actually) or l_j reaches the median of two consecutive elements in $\mathcal{M}^{f_j}$. In the Section 6.2, we will analyze the advantages of this choice.

To sort each $\mathcal{M}^{f_j}$ needs $O(|\mathcal{M}| \log |\mathcal{M}|)$. To search the best score from m_0 to $|\mathcal{M}|$ need at most $O(|\mathcal{M}| \log |\mathcal{N}|)$. In the worst case, to find the best signature in the gap between $x_{m-1}^{f_j}$ and $x_m^{f_j}$ needs to search half of the gap. Since $|S| \leq K$, the total running time is $O(K(|\mathcal{M}| \log |\mathcal{M}| + |\mathcal{M}| \log |\mathcal{N}| + G))$. G is the possible maximum gap among all the fields.

5.3 Signature Pruning

Still we have a set of candidate field and length signatures. Too many length signatures will cause unnecessary false positives, because we try to match any of the length signatures in the detection phase. Therefore, in this final step, we will find an optimal small subset of signature candidates to be the final signature set. Usually, the more signatures we use, the more false positives there are, but with better coverage.

As proved in Section 4.3, to select the optimal small set of signatures in general is NP-Hard. The algorithm proposed here is not to search for global optimum, but find a good solution with bounded false positives and negatives. In the Step 3 algorithm, γ' and γ are parameters and $\gamma' < \gamma$. The Step 3 algorithm has two stages. The first one is the opportunistic stage. We opportunistically find the signatures which can improve at least γ' percent of the initial suspicious pool coverage than the existing signature set without generating any false positive. Usually, γ' is small. If the best approximated signatures for each worm have 0 false positive, the opportunistic stage can help improve the true positives even when adversaries are present. Then, we use a similar process to find other signatures with marginal improvement requirement γ.

Calculating $|\mathcal{M}_s|$ takes $O(\log |\mathcal{M}|)$, and thus finding the signature with maximum coverage takes $O(K \log |\mathcal{M}|)$. Furthermore, removing samples matched by signature s takes $O(|\mathcal{M}|)$. Therefore, final running time for Step 3 algorithm can be bounded by $O(K(K \log |\mathcal{M}| + |\mathcal{M}|))$.

With our three-step algorithm, we guarantee low false positives and false negatives on the generated signatures for buffer-overflow worms. For non-buffer-overflow worms, the algorithm will output an empty set finding no signatures to meet the minimal requirements on false positives and false negatives.

6 Attack Resilience Analysis

We presented the length signature generating algorithm and its performance analysis in the previous section. In this section, we analyze and prove attack resilience of our algorithm, *i.e.*, the quality of signatures generated (evaluated by false negatives and false positives) when attackers launch attacks to try to confuse and evade the LESG system. In particular, attackers may deliberately inject some noises into the suspicious pool to fool LESG.

6.1 Worst Case Performance Bounds

Note that the noisy length signature generation problem (NLBSG) is a NP-Hard problem and even *the global optimum solution* due to the limited input size can be different from the ground truth signature L_B as defined in Section 4.2. The signatures we generated are *approximated signatures*. In the Step 1 and Step 2 algorithms,

Attackers can fully craft the worms and	The signature B' has zero false negative and	
	zero false positive	non-zero false positive
can craft noises	Theorem 2	Theorem 3
cannot craft noises	Theorem 4	Theorem 5

Table 3: Worst cases with different assumptions

we always select the field f_B in the signature candidate set if the worm coverage is larger than COV_0. But in our algorithm, we might not get the optimal length L_B, instead we get L'_B. We denote the signature $B' = (f_B, L'_B)$. We tend to choose a more conservative signature than the ground truth signature B. Therefore $\mathrm{FN}_{\{B'\}} = 0$ and $\mathrm{FP}_{\{B'\}} \leq \mathrm{FP}_0$.

For most cases, the distributions of normal flows and worm flows are well apart and there is a noticeable gap between the two distributions, then we will get $\mathrm{FP}_{\{B'\}} = 0$ which has the same power as the ground truth signature. Without adversaries, our algorithm will output the signature B', we call it the *best approximated signature* because it has the tightest bound to the corresponding ground truth signature when compared with signatures generated with adversaries. Then with different adversary models and depending on whether the normal and worm flow length distributions have a noticeable gap, our algorithm will output different approximated signatures with different attack resilience bounds. In this section, we prove these bounds when compared with the ground truth signatures.

Let $\mathcal{M}^1$ be the set of true worm flows in $\mathcal{M}$ and let $\mathcal{M}^2 = \mathcal{M} - \mathcal{M}^1$, which is all the noise in the $\mathcal{M}$. Let the fraction of worm flows in $\mathcal{M}$ be α, *i.e.* $\frac{|\mathcal{M}^1|}{|\mathcal{M}|} = \alpha$. For simplicity, in the Step 3 algorithm, we denote the loop of finding zero false positive signatures as $LOOP_1$ and the loop of finding non-zero false positive signatures as $LOOP_2$. Except Theorem 2, the proofs of all the following theorems can be found in Appendix A.

6.1.1 Performance Bounds with Crafted Noises

In Theorems 2 and 3, we prove the worse case performance bounds of our system under the deliberate noise injection attacks, *i.e.*, with crafted noises. This is the worst case. The attackers not only fully craft the worms but also inject the crafted noises. The difference between Theorem 2 and Theorem 3 is that Theorem 2 assumes the length distributions of normal flows and worm flows are well apart which is the most common case in reality. Theorem 3 consider even more general cases, which the length distributions of normal flows and worm flows can have overlap.

Theorem 2. *If the best approximated signature has no false negative and no false positive, the three step algorithm outputs a signature set* Ω *such that* $\mathrm{FN}_\Omega < \frac{\gamma'}{\alpha}$ *and* $\mathrm{FP}_\Omega \leq \mathrm{FP}_0 \cdot \lfloor \frac{1-\alpha}{\gamma} \rfloor$.

Proof. Let the best approximated signature be s. $\frac{|\mathcal{M}^1_{\{s\}}|}{|\mathcal{M}^1|} = 1$ and $\mathrm{FP}_{\{s\}} = 0$. Let the signature set we find in $LOOP_1$ be Ω_1. Let the signature set found in $LOOP_2$ be $\Omega_2 = \Omega - \Omega_1$.

After $LOOP_1$, the residue of true worm samples $|R| < \gamma' \cdot |\mathcal{M}|$. If $|R| \geq \gamma' \cdot |\mathcal{M}|$, s will taken as the output. Then there is no true worm samples left. Therefore $|R| < \gamma' \cdot |\mathcal{M}|$.

Therefore, $|\mathcal{M}^1_\Omega| \geq |\mathcal{M}^1_{\Omega_1}| = |\mathcal{M}^1 - R| = |\mathcal{M}^1| - |R| > |\mathcal{M}^1| - \gamma' \cdot |\mathcal{M}|$. Since $\frac{|\mathcal{M}^1|}{|\mathcal{M}|} = \alpha$, $|\mathcal{M}^1_\Omega| > |\mathcal{M}^1| - \gamma' \cdot |\mathcal{M}| = |\mathcal{M}^1| - \gamma' \cdot \frac{|\mathcal{M}^1|}{\alpha} = (1 - \frac{\gamma'}{\alpha}) \cdot |\mathcal{M}^1|$. Hence, $\frac{|\mathcal{M}^1_\Omega|}{|\mathcal{M}^1|} > 1 - \frac{\gamma'}{\alpha}$. Therefore, $\mathrm{FN}_\Omega < \frac{\gamma'}{\alpha}$.

Suppose the first output signature in $LOOP_1$ is s'; then $\frac{|\mathcal{M}_{\{s'\}}|}{|\mathcal{M}|} \geq \frac{|\mathcal{M}_{\{s\}}|}{|\mathcal{M}|} = \alpha$. Therefore after $LOOP_1$, the remaining suspicious pool size $|\mathcal{M}'| \leq (1 - \alpha) \cdot |\mathcal{M}|$.

Since $\mathrm{FP}_{\Omega_1} = 0$, we have $\mathrm{FP}_\Omega = \mathrm{FP}_{\Omega_2}$. Since in $LOOP_2$ each iteration needs to improve coverage by γ, there at most is $\lfloor \frac{|\mathcal{M}'|}{\gamma \cdot |\mathcal{M}|} \rfloor \leq \lfloor \frac{(1-\alpha) \cdot |\mathcal{M}|}{\gamma \cdot |\mathcal{M}|} \rfloor = \lfloor \frac{1-\alpha}{\gamma} \rfloor$ iterations. Each iteration may introduce false positive ratio $\mathrm{FP} \leq \mathrm{FP}_0$. Therefore the final false positive ratio is bounded by $\mathrm{FP}_0 \cdot \lfloor \frac{1-\alpha}{\gamma} \rfloor$

□

Theorem 3. *If the best approximated signature has no false negative and the false positive ratio is bounded by* FP_0*, the three-step algorithm outputs a signature set* Ω *such that* $\mathrm{FN}_\Omega < \frac{\gamma}{\alpha}$ *and* $\mathrm{FP}_\Omega = \mathrm{FP}_0 \cdot (\lfloor \frac{1-\alpha}{\gamma} \rfloor + 1)$.

These bounds are still tight as shown in the example of deliberated noise injection attacks in Section 6.1.3. Furthermore, the experimental accuracy results obtained in Section 7.7 are even better than these bounds.

6.1.2 Performance Bounds without Crafted Noises

Since injected crafted noises will slow down the worm propagation, the worm authors might not want to do that. For example, when the noise ratio is 90% (*i.e.*, 90% of traffic from a worm is crafted noises), the worm will propagate at least 10 times slower than before based on worms based on the RCS worm model [12]. For example, the Code Red II may take 140 hours (six days) to comprise all vulnerable machines instead of 14 hours.

Without crafted noises, *i.e.*, the noises are from normal traffic, we are able to prove even tighter performance bounds for our system. Here, the Theorem 4 below assumes the length distributions of normal flows and worm flows are well apart while the Theorem 5 removes this assumption. Both theorems assume the noises in the suspicious pool is randomly sampled from the normal traffic.

Theorem 4. *If the noise in the suspicious pool is normal traffic and not maliciously injected, and if the best approximated signature has no false positives and no false negatives, then the three-step algorithm outputs a signature set* Ω *such that* $\mathrm{FN}_\Omega = 0$ *and* $\mathrm{FP}_\Omega = 0$*; in other words, with no false negative and false positive.*

In this case, the outputted signature set Ω contain the best approximated signature.

Theorem 5. *If the noise in the suspicious pool is normal traffic and not maliciously injected, and if the best approximated signature has no false negative and false positive ratio is bounded by* FP_0*, then the three-step algorithm outputs a signature set* Ω *such that* $\mathrm{FN}_\Omega \leq \mathrm{FP}_0 \cdot \frac{1-\alpha}{\alpha}$ *and* $\mathrm{FP}_\Omega \leq \mathrm{FP}_0$.

The evaluation results in Section 7.2 are consistent with the theorem and are often better than the bounds proved in the theorems.

6.1.3 Discussions

In this section, we discuss some issues related to the attack resilience theorems.

Multiple worms. For single worm cases, the theorems can be directly applied. In the case that multiple worms are in the suspicious pool, for each worm, we treat the other worms as noises, and thus we have the same bound.

Parameter FP_0. From the theorems above, we can tell that FP_0 plays an important role on the bound. We have the following observations for its value. Usually given a standard protocol, a popular implementation of peer/server should be able to interoperate with various different implementations of peer/clients. Thus, even for a server implementation with a buffer overflow vulnerability, in most cases the normal traffic should not trigger the buffer overflow. Here we assume FP_0 is no larger than 0.1% and we conservatively set it to be 0.1%, *i.e.*, the server should be able to handle 1000 normal requests without crash (buffer overflow triggered). This is equivalent to a server handle six requests per hour and does not crash for a week. We believe this is reasonable for most popular implementations of a protocol.

Assumptions for theorems on attack resilience. There are two general assumptions for all the theorems above. First, there is little or no overlap for the input length of vulnerable fields between the normal traffic and the worms. This is discussed in Section 4.2 and also validated in our experiments. Secondly, the attacker cannot change the field length distribution of normal traffic which is also generally true. Compared with the recent Hamsa system [43], we have less assumptions and allow crafted noises.

6.2 Resilience Against the Evading Attacks

In this section, we discuss the resilience of our schemes against several recently proposed attacks [20–22].

Deliberate noise injection attack In [19], deliberate noise injection is proposed to mislead the worm signature generator. Most other existing signature generators suffer under this attack. However, even with this attack, our approach can perform reasonably well, especially in the case when the best approximated signature with zero false positive exists. For example if $\gamma' = 1\%$ and $\gamma = 5\%$, even with 90% crafted noise, in most cases, the false negative rate can be bound as 10% and the false positive rate, 1.8%. Also, the experiments results in Section 7.7 is much better than the bound: Under the deliberate noise injection attack with 90% craft noises, the false negative rate is 6.3% and false positive rate is 0.14%. To the best of our knowledge, we are the *first* network-based approach that can achieve this performance.

There are several different attacks proposed in Paragraph [20]. Among them, the *suspicious pool poisoning attack* is similar to the deliberate noise injection attack. Next, we discuss other attacks.

Randomized red herring attack or coincidental attack is to inject unnecessary tokens to the content based approaches with a probability model, so that these tokens are highly likely to be included in the signatures to produce more false negatives. A similar attack can be proposed to our approach. However, that requires the attackers to use the "don't care" fields, the fields which can be manipulated without influencing the worm execution. Unlike the content-based signature generation approaches with which attackers can inject as many tokens as they want, there may be zero or only a small number of such "don't care" fields in a protocol, so the attack may not be applicable. Moreover we use a signature set. When any signature in this set matches the sample, we label the sample as a worm. This is more resilient than using the whole set as a signature.

Dropped red herring attack includes some tokens in the beginning of the worm spread and drops those tokens in later propagation of the worm. Again, a similar attack can be proposed for our approach. However, there are several problems as well as countermeasures for such attacks. First, this attack also requires "don't care" fields. Secondly, we can potentially still detect the worm with any disjunction in the signature set instead of using the conjunction. Thirdly, this attack is hard to implement because it requires the worm to dynamically change itself with synchronized actions. Fourthly, there are some dynamic update problems for signature change and signature regeneration. Since our signature generation is fast, it can alleviate the damage by this attack.

Moreover, there is another similar attack which can be designed specially for length-based signatures. We call it *length dropping attack*. Since the attackers have to inject an input longer than the buffer length L_B, they can inject a long input L at the beginning and gradually decrease the length by ΔL in each run of infection until L_B. However, since if there is a gap between L and L_B, in our design we choose the signature length be $l_j = \frac{x_{m-1}^{f_j} + x_m^{f_j}}{2}$, and the $x_{m-1}^{f_j}$ is comparable to L_B and the $x_m^{f_j}$ is comparable to L. In other words, we will choose the median of L and L_B. Therefore, even when this attack is launched, we only need regenerate the length signature $O(\log(L_1 - L_B))$ times where L_1 is the initial length that the attackers use.

Innocuous pool poisoning is to poison the normal traffic pool. However, in general, this is very hard. First, the amount of normal traffic is huge, even to poison 1% is hard. Second, using the random selection policy of normal traffic [14], it is very hard for attackers to poison the traffic in the right time to have an effective evasion during the worm breakout.

In [21], Simon *et al.*propose two types of *allergy attacks*. The type I attack makes the IDS to generate signatures which can deny current normal traffic. The type II attack makes the IDS generate signatures which can deny future normal traffic. The type I allergy attack does not work for our approach because we check the false positive to the normal traffic. The type II attack may work in theory, but in practice it is very hard to happen. The contents of future traffic may change a lot than that of the current normal traffic, but the length profile of fields in the protocol will still remain stable. Therefore it is hard to find such a case. Even if there is such a case, it is very hard for the attack to predict.

The *blending attacks* [22] cannot work for our approach because the worm has to use a longer-than-normal input for the vulnerable field and they cannot mimic the normal traffic.

Protocol	DNS	SNMP	$SNMP_{trap}$	FTP_1	FTP_2	FTP_3	SMTP	HTTP
Bugtraq ID	2302	1901	12283	16370	9675	20497	19885	2880
ground truth (fieldID,BufLen)	(2,493)	(6,256)	(7,512)	(1, 228)	(11,419)	(33, 4104)	(3, unknown)	(6, 240)
signature related field length	fixed	variable	variable	variable	variable	variable	variable	variable

Table 4: The summary of worms

7 Evaluation

We implemented the protocol parsing using Perl scripts with BINPAC and Bro, as mentioned in Section 3.1, and implemented the LESG signature generator in *MATLAB*.

7.1 Methodology

We constructed the traffic of eight worms based on real-world exploits, and collected more than 14GB Internet traffic plus 123GB Email SPAM. To test LESG's effectiveness, we used completely different dataset for LESG signature generation (i.e. training dataset) and for signature quality testing (i.e. evaluation dataset). For training dataset, we used a portion of the worm traffic plus some samples from the normal traffic (as noise) to construct the suspicous pool, and a portion of the normal traffic as the normal pool. For evaluation dataset, we used the remaining normal traffic to test the false positives and worm traffic to test false negatives. For attack resilience testing, we tested the performance of our system under deliberate noise injection attack with different noise ratios.

7.1.1 Polymorphic Worm Workload

To evaluate our $LESG$ system, We created eight polymorphic worms based on real-world vulnerabilities and exploits from `securityfocus.com`, as shown in Table 4, by modifying the real exploits to make them polymorphic. The eight worms are with six different protocols, DNS, SNMPv1, $SNMPv1_{trap}$, FTP, SMTP and HTTP. Since the original exploit codes are not polymorphic and the field lengths are fixed, we modified them as follows: for the exploit unrelated fields, *i.e.*"don't care" fields, we randomly chose the lengths with the same distribution as those in normal traffic; for the signature related fields, the lengths in the original exploit codes are longer than the buffer lengths in most cases, so we used these values as upper bound in the worms, and the hidden buffer length or a larger value that we believe is necessary to exploit the vulnerability as the lower bound (specified by the row "ground truth" in Table 4); moreover, for some exploits that have rigid exploit condition, we kept the fixed length. In the Table 4, the row titled "signature related field length" specifies whether the overflowing field length is fixed or not. For the vulnerability that we cannot find the ground truth by searching literatures, we indicate so as "unknown".

The detailed descriptions of the worms we created are as follows.

DNS worm. It's a variance of the lion worm that attacks a vulnerability of BIND 8, the most popular DNS server. The exploit code constructs a UDP DNS message with QUESTION section whose length is 493 bytes and hard to be variable.

SNMP worm. It attacks a vulnerability in the NAI sniffer agent. The vulnerable buffer is 256 bytes long and stores the data transferred in field ObjectSyntax.

SNMP Trap worm. The worm targets Mnet Soft Factory NodeManager Professional. When it processes SNMP Trap messages, it allocates a buffer of 512 bytes to store the data transferred in field ObjectSyntax.

FTP worm I. It exploits a vulnerability in the Sami FTP Server. The content of the USER command must be longer than 228 bytes to overflow the buffer storing it.

FTP worm II. It targets a popular desktop FTP server, Serv-U. The content of the SITE CHMOD command plus a path name is stored in a buffer which is 419 bytes long.

FTP worm III. It targets BulletProof FTP Client. The content of FTP reply code 220 must be longer than 4104 bytes.

SMTP worm. This vulnerability resides in the RCPT TO command of the Ipswitch IMail Server.

HTTP worm. It exploits the IIS vulnerability also attacked by a famous worm Codered. The difference is we varied the length of our created worm, while Codered has fixed length.

7.1.2 Normal Traffic Data

Number of Fields	Normal pool			Evaluation dataset		
	Bytes	Flows	Hours	Bytes	Flows	Hours
DNS: 14	120M	320K	21	960M	4.4M	120
SNMP: 10	12M	13K	20	282M	77K	120
SNMP_t: 11	21M	16K	72	67M	54K	218
FTP: 60	2.7G	66K	14	10G	373K	37
SMTP: 13	840M	210K	24	122G	31M	744
HTTP: 7	2G	77K	7	11G	360K	40

Table 5: Dataset summary for evaluation

The traffic traces were collected at the two gigabit links of the gateway routers at Tsinghua University campus network in China, in June 21 - 30, 2006. All traffic of Tsinghua University to/from DNS, SNMPv1 Trap, SNMPv1, HTTP and FTP control channel are collected, without using any form of sampling. We used another 123GB SPAM dataset from some open relay servers in a research organization at US for the SMTP. The datasets are summarized in Table 5. Since SNMPv1 Trap message is sent to port 162 and its format is different from other types of messages, we treat SNMPv1 Trap as a protocol separate from SNMPv1 on port 161. Also note that for evaluation purpose, in our prototype system we only parsed the GET request for HTTP, which has the same effect as a complete parsing, because the worm is only related to GET request. The traces are checked by the Bro IDS system to make sure that the traces are normal traffic.

7.1.3 Experiment Settings

In the Step 1 algorithm, we set $\text{FP}_0 = 0.1\%$ and $\text{COV}_0 = 1\%$. The score function in Step 2 is $Score(\text{COV}, \text{FP}) = (1/log\text{FP} + 1) * \text{COV}$, which works well in practice. The basic requirement of a score function is that the score should be monotonically increasing with COV and decreasing with FP. This function has another merit that a large FP (eg. $\text{FP} \in [10^{-2}, 10^{-1}]$) affects the score greater than a much smaller FP (eg. $\text{FP} \in [10^{-5}, 10^{-4}]$) does. In Step 3, we choose $\gamma' = 1\%$ and $\gamma = 5\%$, indicating that we focus on the worms that cover more than 1% of the suspicious pool.

All experiments were conducted on a PC with a 3.0GHz Intel Xeon running Linux kernel 2.6.11.

7.2 Signature Generation for A Single Worm with Noise

We evaluated the accuracy of LESG with presence of noise. The noise is the flows randomly sampled from normal traffic, and mixed with worm samples to compose the suspicious pool. We chose DNS, SNMP, SNMP_{trap}, SMTP and HTTP protocols to demonstrate the cases of single worm with noise. For HTTP we also tested our algorithm against Codered worm.

For each protocol, we tested the suspicious pool size of 50, 100, 200 and 500, and at each size we changed the noise ratio from 0 to 80% increasing 10% in each test. After signature generation, we matched the signatures against another 2000 samples of worms and evaluation set of normal traffic to test the sensitivity and accuracy.

Table 6 shows the range of the signatures we generated and their accuracy. Tr. FN/FP denotes training false negatives and false positives in the training data. Ev. FN/FP denotes the evaluation false negatives and false positives in evaluation data set. Under all the pool sizes and noise ratios, the same signature fields are generated. Because the size of suspicious pool is limited, the signature length varies in different tests. We checked these signatures against the evaluation datasets, and they all have excellent false negative and false positive ratio. It may be noticed that generated signature lengths are smaller than the true buffer length, because the length in normal flows are usually much smaller than the buffer length, which is reasonable since the buffer length is designed to be longer than the longest possible normal requests.

Worm	Signatures (ID,length)	Tr. FN	Tr. FP	Ev. FN	Ev. FP
DNS	(2, 284∼296)	0	0	0	0
SNMP	(6, 133∼238)	0	0	0	0
$SNMP_t$	(7, 304∼314)	0	0	0	0
SMTP	(3, 109∼112)	0	0	0	10^{-5}
FTP	(1, 128∼169) (11, 262∼300) (33, 2109∼2121)	0	0	0	0
HTTP	(6, 239∼240)	0	0	0∼1%	10^{-4}
CodeRed	(6, 339)	0	0	0	10^{-5}

Table 6: Signatures and accuracy under different pool size and noise

7.3 Signature Generation for Multiple Worms with Noise

We also evaluated the case of multiple worms with noise using the FTP protocol. We have three FTP worms in total. we tested the suspicious pool size of 50, 100, 200 and 500, and at each size we change the noise ratio from 0 to 70% increasing 10% in each test.The result is also shown in Table 6.

7.4 Evaluation of Different Stages of the LESG Algorithm

The LESG algorithm has three steps, and we evaluated the effectiveness of each step. Table 7 illustrates the results of each step for the DNS worm, with a suspicious pool of size 100 and noise ratio 50%. Table 7 shows that the false positive rate is largely decreased by refining each signature length in Step 2. And comparing with Table 4, we can see that in Step 3, the best and most accurate signature is selected, further decreasing the false positives.

	Signature	Tr. FN	Tr. FP
Step 1	{(1,62), (2,66), (3,2), (4,15), (5,28), (6,47), (10,99),(11,2)}	0	0.32%
Step 2	{(1,68), (2,296), (3,21), (4, 99), (5,333), (6,543), (10,111), (11,2)}	0	0.15%
Step 3	{(2, 296)}	0	0

Table 7: Result of each step for the DNS worm

7.5 Pool Size Requirement

We tested the accuracy of our algorithm when only a small suspicious pool is available. We chose suspicious pools of size 10 with noise ratio 20%, and size 20 with noise ratio 50%. All the tests generated signatures within the range presented in Table 6.

We further did similar tests for the DNS worm using different normal pool size 5K, 10K, 20K, and 50K. And we found our approach is not sensitive to the size of normal pool either.

7.6 Speed and Memory Consumption Results

	Normal pool (Bytes/Flows) (Bytes/Flows)	Protocol parsing (secs)	Signature generation (in different pool size) (secs)			
			50	100	200	500
DNS	120M/320K	58	2.1	3.6	9.4	18
SNMP	12M/13K	8	0.08	0.09	0.15	0.32
$SNMP_t$	21M/16K	4	0.12	0.24	0.37	0.88
FTP	2.7G/66K	95	0.20	0.29	0.54	1.20
SMTP	836M/210K	50	0.47	1.30	1.84	3.36

Table 8: Speed of protocol parsing and signature generation

We evaluated the parsing speed by using Bro and BINPAC, and the speed of our signature generation

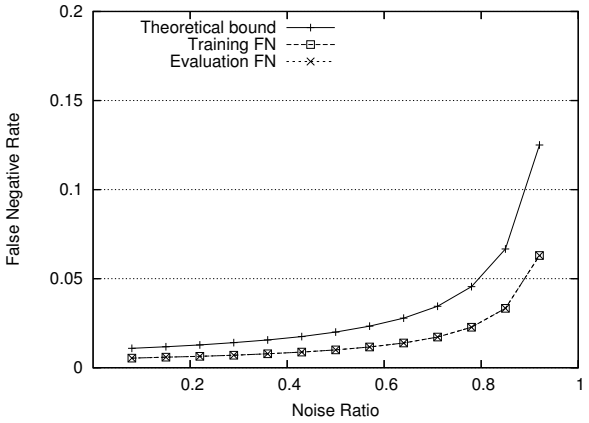

Figure 6: False negative rate with different noise ratio

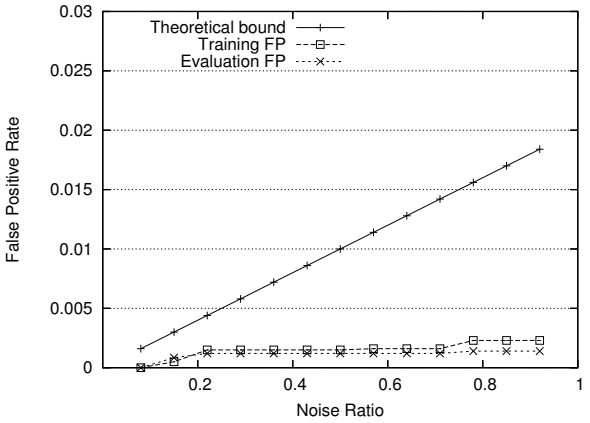

Figure 7: False positive rate with different noise ratio

algorithm. Since HTTP was not completely parsed, we only provide the result of the other five protocols. Table 8 shows that the speed of signature generation algorithm are quite fast, though the speed is influenced by the size of the suspicious pool and the normal pool. The protocol parsing for normal pool can be done offline. We can run the process once a while (e.g. several hours). And these datasets were collected over a 20-hour+ period. For the suspicious pool, since it is much smaller than the normal pool, the protocol parsing can be done very quickly. Moreover, as mentioned in [44], the BINPAC compiler can be built with parallel hardware platforms, which makes it much faster.

Normal pool size		Suspicious pool size		
		100	200	500
DNS (14 fields)	50K	5.64MB	5.66MB	5.71MB
	100K	11.26MB	11.28MB	11.33MB
FTP (60 fields)	50K	8.43MB	8.45MB	8.53MB
	100K	16.83MB	16.85MB	16.93MB

Table 9: Memory usage of the algorithm

The memory usage of the signature generation algorithm implemented in *Matlab* was evaluated under different pool sizes, shown in Table 9. The memory usage is proportional to the normal pool size and the number of fields.

7.7 Performance under Deliberate Noise Injection Attacks

In [19], two deliberate noise injection attacks are implemented targeting the token based signature generation systems, such as Polygraph. Inspired by their work, we implemented a similar attack to the $LESG$ system. In the attack we implemented 1) the attackers know all the parameters used in our system and optimize the attack against them; 2) the attackers can obtain certain normal traffic samples, so that they can estimate field length distributions.

We demonstrated this attack by modifying the Lion worm of the DNS protocol. There are 14 fields of the DNS protocol. Only one field f_B has to be long enough to overflow the buffer, which cannot be controlled by attackers. The attackers can use the other 13 fields to craft arbitrary noises.

In the experiment we simulated the situation that the attacker capture the normal traffic with 100K flows to optimize the attack. To make sure the attacker's normal traffic has similar length distributions as the training normal pool we used. In our experiment we randomly permutated the 320K DNS normal flows shown in Table 5, and divide into the 100K flow pool for attackers and the 220K flow normal pool for the $LESG$ system.

In this experiment, we assume all the noise are deliberately injected. We test the noise ratio from 8% to 92% increasing 7% in each test. We use suspicious pool size of 200. For the Lion worm, the best approximate signature has no false positive. The proof of Theorem 2 shows that the false negatives should be less than γ'

portion of the suspicious pool. Therefore, at most $200 \times \gamma' - 1 = 200 \times 0.01 - 1 = 1$ false negatives can be generated. Under a given noise ratio, among the 13 fields, we search all the possible combinations and choose the optimal set of fields to increase the false positives. Then we choose one of the remaining fields to increase the false negatives.

we use another 2000 worms to test evaluation false negatives and the 4.4M DNS flows shown in Table 5 to test the evaluation false positives. The results is shown in Figure 6 and Figure 7. In training false negatives and the evaluation false negatives are very close, so the two line collide each other. From the results we know, even with 90% deliberate injected noise, our system still only has 6.3% false negative and 0.14% false positive. This indicates it is quite hard for attackers to increase the false positives. The reason behind this is that the worst case bound happens when each field can introduce false positives in a mutual exclusive way, which is not almost possible in practice.

8 Discussions of Practical Issues

Speed of Length Based Signature Matching The operation of length-based signature matching has two steps: protocol parsing of packets and field length comparison with the signatures. The latter is trivial. The major overhead is for protocol parsing. Currently, the Bro and BINPAC based parsing can achieve $50 \sim 200$ Mbps. As mentioned in [44], with parallel hardware platform support, BINPAC may achieve $1 \sim 10$ Gbps. On the commercial products side, Radware's security switch on ASIC based network processor can operate at 3 Gbps link with protocol parsing capability [45]. Therefore, with hardware support, the whole length based signature matching can be done very fast, which is comparable to current speed of pattern-based (string) signature matching techniques widely used in IDSs.

Relationship Between Fields and Vulnerable Buffers The main assumption of length based signatures is that there is a direct mapping between variable length fields and vulnerable buffers. In addition to the vulnerabilities shown in the evaluation section, we further checked 11 more buffer overflow vulnerabilities from `securityfocus.com`. We found that the assumption hold for all cases except one. Next, we will first examine the normal cases and then check the special one.

In Section 4.1 we show that the consecutive fields can be combined together to a *compound field*. For the variable length fields which cannot be further decomposed, we call them *simple fields*. We found in 13 cases (out of the total of 18 cases that we examined) the field mapped to the vulnerable buffer is a simple field while in 3 cases it is a compound field. There is one case we found that two simple fields, which cannot be combined to a compound field, are mapped to one vulnerable buffer. Therefore, either of the two fields can cause the buffer overflow to happen. In all these cases, we can get the accurate length-based signatures. However, we did find one case (again, 1 out of 18 cases) which does not have length-based signatures. It is a buffer overflow vulnerability present in versions of wu-ftpd 2.5 and below. The vulnerable buffer corresponds to the path of the directory. So if a very deep path is created by continuously making new directories recursively, the buffer will eventually be overflowed. From the protocol messages of the FTP, only a set of MKD (mkdir) commands can be seen and the length of each directory could be normal. Therefore no length-based signatures exist.

9 Conclusions

In this paper, we proposed a novel network-based approach using protocol semantic information to generate length-based signatures for buffer overflow worms. This is the first attempt to generate vulnerability based signatures at network level. Our approach has good attack resilience guarantees even under deliberate noise injection attacks. We further show that our approach is fast and accurate with small memory consumption through evaluation based on real-world vulnerabilities, exploit codes and real network traffic.

References

[1] Z. Liang and R. Sekar, "Automatic generation of buffer overflow attack signatures: An approach based on program behavior models," in *Proc. of Computer Security Applications Conference (ACSAC)*, 2005.

[2] E. Mars and J. D. Jansky, "Email defense industry statistics," http://www.mxlogic.com/PDFs/IndustryStats.2.28.04.pdf.

[3] Marty Roesch, "Snort: The lightweight network intrusion detection system," 2001, http://www.snort.org/.

[4] Vern Paxson, "Bro: A system for detecting network intruders in real-time," *Computer Networks*, vol. 31, 1999.

[5] C. Kreibich and J. Crowcroft, "Honeycomb - creating intrusion detection signatures using honeypots," in *Proc. of the Workshop on Hot Topics in Networks (HotNets)*, 2003.

[6] S. Singh, C. Estan, et al., "Automated worm fingerprinting," in *Proc. of USENIX OSDI*, 2004.

[7] H. Kim and B. Karp, "Autograph: Toward automated, distributed worm signature detection," in *Proc. of USENIX Security Symposium*, 2004.

[8] Z. Liang and R. Sekar, "Fast and automated generation of attack signatures: A basis for building self-protecting servers," in *Proc. of ACM CCS*, 2005.

[9] X. Wang, Z. Li, J. Xu, M. Reiter, C. Kil, and J. Choi, "Packet vaccine: Black-box exploit detection and signature generation," in *Proc. of ACM CCS*, 2006.

[10] D. Brumley, J. Newsome, D. Song, H. Wang, and S. Jha, "Towards automatic generation of vulnerability-based signatures," in *Proc. of IEEE Security and Privacy Symposium*, 2006.

[11] David Moore, Vern Paxson, Stefan Savage, Colleen Shannon, Stuart Staniford, and Nicholas Weaver, "The spread of the Sapphire/Slammer worm," http://www.caida.org, 2003.

[12] Stuart Staniford, Vern Paxson, and Nicholas Weaver, "How to own the Internet in your spare time," in *Proceedings of the 11th USENIX Security Symposium*, 2002.

[13] S. Staniford, D. Moore, V. Paxson, and N. Weaver, "The top speed of flash worms," in *Proc. of ACM CCS WORM Workshop*, 2004.

[14] Z. Li, M. Sanghi, Y. Chen, M. Kao, and B. Chavez, "Fast signature generation for zero-day polymorphic worms with provable attack resilience," in *Proc. of IEEE Security and Privacy Symposium*, 2006.

[15] J. Newsome, B. Karp, and D. Song, "Polygraph: Automatically generating signatures for polymorphic worms," in *Proc. of IEEE Security and Privacy Symposium*, 2005.

[16] Yong Tang and Shigang Chen, "Defending against internet worms: A signature-based approach," in *Proc. of IEEE Infocom*, 2003.

[17] James Newsome and Dawn Song, "Dynamic taint analysis for automatic detection, analysis, and signature generation of exploits on commodity software," in *Proc. of NDSS*, 2005.

[18] J. R. Crandall, Z. Su, and S. F. Wu, "On deriving unknown vulnerabilities from zeroday polymorphic and metamorphic worm exploits," in *Proc. of ACM CCS*, 2005.

[19] R. Perdisci, D. Dagon, W. Lee, et al., "Misleading worm signature generators using deliberate noise injection," in *Proc. of IEEE Security and Privacy Symposium*, 2006.

[20] James Newsome, Brad Karp, and Dawn Song, "Paragraph: Thwarting signature learning by training maliciously," in *Proc. of International Symposium On Recent Advances In Intrusion Detection (RAID)*, 2006.

[21] Simon P. Chuang and Aloysius K. Mok, "Allergy attack against automatic signature generation," in *Proc. of International Symposium On Recent Advances In Intrusion Detection (RAID)*, 2006.

[22] Prahlad Fogla, Monirul Sharif, Roberto Perdisci, Oleg Kolesnikov, and Wenke Lee, "Polymorphic blending attacks," in *Proc. of USENIX Security Symposium*, 2006.

[23] V. Yegneswaran, J. Giffin, P. Barford, and S. Jha, "An architecture for generating semantic-aware signatures," in *Proc. of USENIX Security Symposium*, 2005.

[24] Christopher Kruegel, Engin Kirda, et al., "Polymorphic worm detection using structural information of executables," in *Proc. of Recent Advances in Intrusion Detection (RAID)*, 2005.

[25] Packeteer, "Solutions for Malicious Applications," http://www.packeteer.com/prod-sol/solutions/dos.cfm.

[26] K. Wang and S. J. Stolfo, "Anomalous payload-based network intrusion detection," in *Proc. of Recent Advances in Intrusion Detection (RAID)*, 2004.

[27] K. Wang, G. Cretu, and S. J. Stolfo, "Anomalous payload-based worm detection and signature generation," in *Proc. of Recent Advances in Intrusion Detection (RAID)*, 2005.

[28] R. Vargiya and P. Chan, "Boundary detection in tokenizing network application payload for anomaly detection," in *Proc. of ICDM Workshop on Data Mining for Computer Security (DMSEC)*, 2003.

[29] M. Cost, J. Crowcroft, M. Castro, A. Rowstron, L. Zhou, L. Zhang, and P. Barham, "Vigilante: End-to-end containment of internet worms," in *Proc. of ACM Symposium on Operating System Principles (SOSP)*, 2005.

[30] A. Pasupulati et al., "Buttercup: On network-based detection of polymorphic buffer overflow vulnerabilities," in *Proc. of IEEE/IFIP Network Operations and Management Symposium (NOMS)*, 2004.

[31] F. Hsu and T. Chiueh, "Ctcp: A centralized TCP/IP architecture for networking security," in *Proc. of ACSAC*, 2004.

[32] X. Wang et al., "Sigfree: A signature-free buffer overflow attack blocker," in *Proc. of USENIX Security Symposium*, 2006.

[33] R. Chinchani and E. Berg, "A fast static analysis approach to detect exploit code inside network flows," in *Proc. of Recent Advances in Intrusion Detection (RAID)*, 2005.

[34] Critical Solutions Ltd., "Critical TAPs: Ethernet splitters designed for IDS," `http://www.criticaltap.com`.

[35] V. Yegneswaran, P. Barford, and D. Plonka, "On the design and use of internet sinks for network abuse monitoring," in *Proc. of RAID*, 2004.

[36] Michael Bailey, Evan Cooke, Farnam Jahanian, Jose Nazario, and David Watson, "The internet motion sensor: A distributed blackhole monitoring system," in *Proc. of NDSS*, 2005.

[37] W. Cui, V. Paxson, and N. Weaver, "Gq: Realizing a system to catch worms in a quarter million places," Tech. Rep. TR-06-004, ICSI, 2006.

[38] Y. Gao, Z. Li, and Y. Chen, "A dos resilient flow-level intrusion detection approach for high-speed networks," in *Proc. of the IEEE International Conference on Distributed Computing Systems (ICDCS)*, 2006.

[39] R. Pang, V. Paxson, R. Sommer, and L. Peterson, "binpac: A yacc for writing application protocol parsers," in *Proc. of ACM/USENIX Internet Measurement Conference*, 2006.

[40] "The ethereal network analyzer," `http://www.ethereal.com/`.

[41] P. V. Mockapetris and K. Dunlap, "Development of the domain name system," in *Proceedings of ACM SIGCOMM '88*. ACM, August 1988.

[42] Staal A. Vinterbo, "Maximum k-intersection, edge labeled multigraph max capacity k-path, and max factor k-gcd are all NP-hard," Tech. Rep., Decision Systems Group,Harvard Medical School, 2002.

[43] Z. Li, M. Sanghi, Y. Chen, M. Kao, , and B. Chavez, "Hamsa: Fast signature generation for zero-day polymorphic worms with provable attack resilience," in *Proc. of IEEE Symposium on Security and Privacy*, 2006.

[44] V. Paxson, K. Asanovic, S. Dharmapurikar, J. Lockwood, R. Pang, R. Sommer, and N. Weaver, "Rethinking hardware support for network analysis and intrusion prevention," in *Proc. of USENIX Hot Security*, 2006.

[45] Radware Inc., "Introducing 1000X Security Switching," `http://www.radware.com/content/products/application_switches/ss/default%.asp`.

A Proofs

Proofs of Performance Bounds with Crafted Noises

Lemma 1. *If the best approximated signature has zero false positive, in any iteration of $LOOP_1$, the coverage of the output signature in that iteration $\frac{|\mathcal{M}_{i\{s'_i\}}|}{|\mathcal{M}|}$ should be larger than or equal the coverage of the remaining true worms $\frac{|\mathcal{M}_i^1|}{|\mathcal{M}|}$.*

Proof. Let the output signature during the ith iteration be s'_i. Let the best approximated signature be s. Let the suspicious pool right before ith iteration be $\mathcal{M}_i$. Let H_i denote the statement that $|\mathcal{M}_{i\{s'_i\}}|/|\mathcal{M}| \geq \frac{|\mathcal{M}_i^1|}{|\mathcal{M}|} = \alpha'$ where α' is the remaining coverage of the true worms.

If the approximated signature s has been output, $\alpha' = 0$, so H_i is true. Otherwise, if H_i is not true, $|\mathcal{M}_{i\{s'_i\}}| < |\mathcal{M}_i^1|$. Since $|\mathcal{M}_i^1| \leq |\mathcal{M}_{is}|$, s is better than s'_i which cannot happen. Therefore H_i is true in all cases. □

Lemma 2. *If the best approximated signature has non-zero false positive, in any iteration of $LOOP_2$, the coverage of the output signature in that iteration $\frac{|\mathcal{M}_{i\{s'\}}|}{|\mathcal{M}|}$ should be larger than or equal the coverage of the remaining true worms $\frac{|\mathcal{M}_i^1|}{|\mathcal{M}|}$.*

Proof Sketch. The proof is the same as that of Lemma 1, so we omit it here. □

Proof of Theorem 3

Proof. Let the best approximated signature be s. $\frac{|\mathcal{M}^1_{\{s\}}|}{|\mathcal{M}^1|} = 1$ and $\mathrm{FP}_{\{s\}} \leq \mathrm{FP}_0$. Let the signature set we find in $LOOP_1$ be Ω_1. Let the signature set found in $LOOP_2$ be $\Omega_2 = \Omega - \Omega_1$.

After $LOOP_2$ the residue of true worm samples $|R| < \gamma \cdot |\mathcal{M}|$. This can be proved with a similar way as in Theorem 2.

Therefore,$|\mathcal{M}^1_\Omega| = |\mathcal{M}^1 - R| = |\mathcal{M}^1| - |R| > |\mathcal{M}^1| - \gamma \cdot |\mathcal{M}|$. Since $\frac{|\mathcal{M}^1|}{|\mathcal{M}|} = \alpha$, Hence $\frac{|\mathcal{M}^1_\Omega|}{|\mathcal{M}^1|} > 1 - \frac{\gamma}{\alpha}$. Therefore, $\mathrm{FN}_\Omega < \frac{\gamma}{\alpha}$.

After $LOOP_1$, let the remaining suspicious pool be $\mathcal{M}'$. $|\mathcal{M}'^2| = |\mathcal{M}^2| - |\mathcal{M}^2_{\Omega_1}|$. Let the signature outputted in the first iteration of $LOOP_2$ be s'. According to Lemma 2, $|\mathcal{M}'_{\{s'\}}| \geq |\mathcal{M}'^1|$. Therefore the size of the total remaining portion of the suspicious pool after the first iteration of $LOOP_2$ is $|\mathcal{M}' - \mathcal{M}'_{\{s'\}}| = |\mathcal{M}'| - |\mathcal{M}'_{\{s'\}}| \leq |\mathcal{M}'| - |\mathcal{M}'^1| =$

$|\mathcal{M}'^2| = |\mathcal{M}^2| - |\mathcal{M}^2_{\Omega_1}|$.

Since in $LOOP_2$ each iteration needs to improve coverage by γ,we have at most $\lfloor \frac{|\mathcal{M}' - \mathcal{M}'_{\{s'\}}|}{\gamma \cdot |\mathcal{M}|} \rfloor \leq \lfloor \frac{|\mathcal{M}^2| - |\mathcal{M}^2_{\Omega_1}|}{\gamma \cdot |\mathcal{M}|} \rfloor \leq \lfloor \frac{|\mathcal{M}^2|}{\gamma \cdot |\mathcal{M}|} \rfloor = \lfloor \frac{1-\alpha}{\gamma} \rfloor$ more iterations. For $LOOP_2$ totally we have at most $\lfloor \frac{1-\alpha}{\gamma} \rfloor + 1$ iterations. Each iteration introduces at most false positive FP_0. Therefore $\mathrm{FP}_\Omega = \mathrm{FP}_{\Omega_2} \leq \mathrm{FP}_0 \cdot (\lfloor \frac{1-\alpha}{\gamma} \rfloor + 1)$. □

Proofs of Performance Bounds without Crafted Noises

Proof of Theorem 4

Proof Sketch. The proof is by reduction from Theorem 5 with $\mathrm{FP}_0 = 0$ □

Proof of Theorem 5

Proof. Let the best approximated signature to be s. Suppose in $LOOP_1$, we find Ω_1. After removing the samples which have already been covered by Ω_1, let the remaining suspicious pool be $\mathcal{M}'$, and let the true worms in it be $\mathcal{M}'^1$, and the noise in it be $\mathcal{M}'^2 = \mathcal{M}' - \mathcal{M}'^1$. Let the $\frac{|\mathcal{M}'^1|}{|\mathcal{M}|} = \alpha'$.

In $LOOP_2$, let the field selected in the first iteration be f'_1, and the corresponding signature to be s'. The attacker might not want we output s; otherwise we will not have any false negatives. Therefore $|\mathcal{M}'_{\{s'\}}| \geq |\mathcal{M}'_{\{s\}}|$. Since the best approximated signature s will cover all the remaining worms, $|\mathcal{M}'_{\{s\}}| \geq \alpha' \cdot |\mathcal{M}|$. Therefore, $|\mathcal{M}'_{\{s'\}}| \geq \alpha' \cdot |\mathcal{M}|$.

Since $\mathrm{FP}_{\{s'\}} \leq \mathrm{FP}_0$ and the distribution of fields in $\mathcal{M}'^2$ is the same as that in $\mathcal{N}$, $|\mathcal{M}'^2_{\{s'\}}| \leq \mathrm{FP}_0 \cdot |\mathcal{M}'^2|$. Since $|\mathcal{M}'^2| \leq |\mathcal{M}^2|$ and $|\mathcal{M}^2| = (1-\alpha) \cdot |\mathcal{M}|$, $|\mathcal{M}'^2| \leq (1-\alpha) \cdot |\mathcal{M}|$. Therefore,$|\mathcal{M}'^2_{\{s'\}}| \leq \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$.

So we have $|\mathcal{M}'^1_{\{s'\}}| = |\mathcal{M}'_{\{s'\}} - \mathcal{M}'^2_{\{s'\}}| = |\mathcal{M}'_{\{s'\}}| - |\mathcal{M}'^2_{\{s'\}}| \geq \alpha' \cdot |\mathcal{M}| - |\mathcal{M}'^2_{\{s'\}}| \geq \alpha' \cdot |\mathcal{M}| - \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$.

Let the remaining suspicious pool at this stage be $\mathcal{M}_r$. Denote the remaining true worm flows as $\mathcal{M}^1_r$ and the remaining noises as $\mathcal{M}^2_r = \mathcal{M}_r - \mathcal{M}^1_r$. Since we know s' has to cover more than $\alpha' \cdot |\mathcal{M}| - \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$ worms, after removing the worms covered by s', the remaining ones $|\mathcal{M}^1_r| \leq \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$. With Step 1 algorithm, for any remaining signature candidates s, $|\mathcal{M}^2_{r\{s\}}| \leq \mathrm{FP}_0 \cdot |\mathcal{M}^2_r|$. Since $|\mathcal{M}^2_r| \leq |\mathcal{M}^2|$, $|\mathcal{M}^2_{r\{s\}}| \leq \mathrm{FP}_0 \cdot |\mathcal{M}^2|$. Since $|\mathcal{M}^2| = (1-\alpha) \cdot |\mathcal{M}|$, $|\mathcal{M}^2_{r\{s\}}| \leq \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$. Therefore, $|\mathcal{M}_{r\{s\}}| = |\mathcal{M}^1_{r\{s\}} + \mathcal{M}^2_{r\{s\}}| = |\mathcal{M}^1_{r\{s\}}| + |\mathcal{M}^2_{r\{s\}}| \leq 2 \cdot \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}| \leq 2 \cdot \mathrm{FP}_0 \cdot |\mathcal{M}|$. Choose parameters to let $\gamma > 2 \cdot \mathrm{FP}_0$, so no other signature candidates can meet the output standard.

Therefore $\Omega = \Omega_1 \cup \{s'\}$. $\mathrm{FP}_{\Omega_1} = 0$, so we have $\mathrm{FP}_\Omega = \mathrm{FP}_{\{s'\}}$. Since the remaining worms are $\mathcal{M}^1_r$, $|\mathcal{M}^1_\Omega| = |\mathcal{M}^1 - \mathcal{M}^1_r| = |\mathcal{M}^1| - |\mathcal{M}^1_r|$. We know $|\mathcal{M}^1_r| \leq \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$, so $|\mathcal{M}^1_\Omega| \geq |\mathcal{M}^1| - \mathrm{FP}_0 \cdot (1-\alpha) \cdot |\mathcal{M}|$. Since $\frac{|\mathcal{M}^1|}{|\mathcal{M}|} = \alpha$, $|\mathcal{M}^1_\Omega| \geq |\mathcal{M}^1| - \frac{\mathrm{FP}_0 \cdot (1-\alpha)}{\alpha} \cdot |\mathcal{M}^1|$. Therefore, $\frac{|\mathcal{M}^1_\Omega|}{|\mathcal{M}^1|} \geq 1 - \mathrm{FP}_0 \cdot \frac{1-\alpha}{\alpha}$.

Therefore, $\mathrm{FN}_\Omega \leq \mathrm{FP}_0 \cdot \frac{1-\alpha}{\alpha}$ and $\mathrm{FP}_\Omega = \mathrm{FP}_{\{s'\}} \leq \mathrm{FP}_0$.

□

Note that what we have proved is for single worm cases, but it is trivial to extend the proof to multiple worms cases. The difference is that multiple worm cases need multiple iterations. Each iteration is for one worm.