SRHU JOURNAL OF ADVANCED RESEARCH

Multi-hop intrusion and connection-chain detection in blockchain: A survey

[1*]Anupama Mishra, [2]Siddhant Rajhans, [3]Neelam Danu, [3]Vivek Katiyar, and [1]Shiv Preet

[1]Department of Computer Science & Engineering, School of Science & Technology, Swami Rama Himalayan University, Dehradun, India
[2]Stevens Institute of Technology, Hoboken, New Jersey 07030
[3]Department of Applied Sciences & Humanities, School of Science & Technology, Swami Rama Himalayan University, Dehradun, India

Keywords

Anonymity, blockchain, connection chain, cyber-attacks, intrusion detection, stepping stone detection

Abstract

A blockchain is a chain of information-bearing blocks that together form a distributed ledger open to every participant in the network. Although the technology is fairly recent and is mostly linked with cryptocurrencies such as Bitcoin, it has not escaped the attention of intruders. To make matters worse, attackers rely on several tricks to cover the trail of an intrusion, one of which is to route their activity through a series of intermediate hosts, known as a stepping stone chain. Stepping Stone Detection (SSD) techniques are meant to uncover such chains. This paper reviews the existing SSD approaches that have proved effective and discusses the strengths and weaknesses of each. It also explains the background and working principles behind stepping stone detection and outlines how networks can be defended against these attacks.

***Corresponding Author: anupamamishra@srhu.edu.in**

Received: 9 July 2026, received in revised form: 19 July 2026, accepted: 26 August 2026
available online: 28 September 2026; Volume: 1; Issue: 3; Pages: 122-138.

Citation: Mishra, A., Rajhans, S., Danu, N., Katiyar, V., Preet, S. 2026. Multi-hop intrusion and connection-chain detection in blockchain: A survey. SRHU J Adv Res. 1(3): 122-138.

DOI: [to be assigned]

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

1. Introduction

Computer usage has changed enormously since the arrival of the Internet and networked systems. The Internet has reshaped how computers are used in business, academia, transport, finance, medicine and many other sectors. Research into newer technologies such as blockchain continues to widen the reach of the Internet and to improve its efficiency (Internet Society, 2020). A defining property of a blockchain is that data, once written into a block that has joined the chain, is extremely hard to alter. Each block holds a set of valid

transactions; these are hashed, organised into a Merkle tree and placed in the block together with the hash of the preceding block. Because every block carries its predecessor's hash, the chain can be followed backwards all the way to the Genesis Block, i.e., the very first block. The network always favours the longest valid chain. If any block is tampered with, its hash changes, so the block after it now holds a hash that no longer matches; that block and every block after it become invalid, which in effect shortens the chain. Since the longest chain is preferred, the network simply falls back to its last stable state.

Transactions on a blockchain take place directly between peers, with no trusted third party in the middle. This is possible because every node keeps its own copy of the ledger, that is, the complete blockchain. When a sender wants to make a transaction, it broadcasts a message to all the nodes, called miners, together with references to its earlier transactions. The miners check that the sender holds a sufficient balance and then race to solve a hard mathematical puzzle in order to append the new transaction to the chain. The puzzle is to find a random value, the nonce, for which the hash of the transaction meets a predefined condition. Once a nonce has been found, the other nodes can verify it with little effort. A transaction confirmed by at least 51% of the miners is then written to the blockchain as a block.

Alongside its many advantages, blockchain also carries vulnerabilities (Mishra et al., 2018; Gupta and Gupta, 2016, 2015). Such weaknesses give attackers an opening to exploit them and to launch attacks on computers and applications connected to the Internet (Shahmehri et al., 2012; Lai and Hsia, 2007). The range of possible attacks is wide and keeps growing, from malware and phishing to denial-of-service (DoS). Whatever the attack, the attacker looks for a safe route that keeps them anonymous while it is under way. One option is an anonymity tool such as a proxy, which was originally built so that users could browse the Internet without disclosing who they are; today attackers misuse proxies for exactly this purpose (Ding et al., 2009). Another option is to place stepping stones on the attack path, and attacks carried out in this way are called stepping stone attacks. Stepping stones are intermediate hosts that the attacker has already compromised and then uses to relay attack commands (Staniford-Chen and Heberlein, 1995; Zhang and Paxson, 2000). An attacker may, for instance, use a payment gateway as a stepping stone to reach a larger target. The attack is launched over a long connection chain built through hosts that were compromised in earlier attempts. Many anonymity techniques now exist that let Internet users conceal their identity from the outside world (Chaum, 1981; Dingledine et al., 2004), and attackers can turn these same techniques to their advantage to stay hidden while still succeeding, which makes tracing them even harder. Because most countries are connected through the Internet, an attacker can set up and use stepping stones located in another country. The hosts of a single chain may therefore be spread over several countries and still be used together against a target in yet another country or continent (Hsiao and Fan, 2009).

Figure 1 illustrates how an attacker uses stepping stone hosts to issue attack commands: instead of attacking directly, the attacker reaches the victim through intermediate hosts that were compromised beforehand

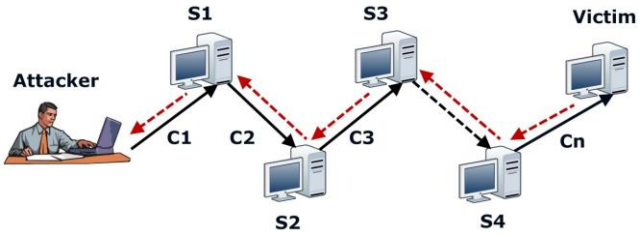


Figure 1: Attack launched through a chain of stepping stone hosts

Web services such as e-commerce, cloud platforms, social networking and email are expanding quickly, both in the number of users and in the range of services offered. People are drawn to them because they make everyday life simpler. The difficulty is that these services hold user information and data, so they need sound security management; without it, both information and money can be lost, and information is at times worth more than money. Even this is not enough, because an attacker may still break through the security measures in place and get into the system to carry out illegal activities. Large organisations such as Google, Yahoo and Amazon have all faced cyber-attacks over the years. At times, service providers even attack their competitors to damage their reputation: when customers see a competitor's weak security, they move to another company that offers the same service with better protection, and the attacking provider gains directly. In most cyber-attacks, the attacker reaches the victim through stepping stones and runs the attack from one of them, usually the last host in the chain, rather than from their own machine. From the attacker's point of view this is the safest option, since it keeps them anonymous while the attack proceeds. The attacker first locates vulnerable computers with scanning techniques, compromises them by exploiting those weaknesses, and then uses them either to run attack commands or to compromise further vulnerable machines. Stepping stones may also be used to consume web services without paying for them, or to deny those services to legitimate users through DoS or DDoS attacks. In most cases the aim is to steal information or to carry out fraudulent transactions for financial gain, which is why online shopping websites and online banking services are frequent targets. There have also been cases where one country used a stepping stone host located in another country to steal secret information from a third, so that the theft appeared to come from the intermediate country rather than from the real attacker (Dingledine et al., 2004; Sauter, n.d.).

Taken together, these points show how serious stepping stone attacks are, and they call for real-time approaches that can detect stepping stones in different attack scenarios, for example with single or multiple measurement points, and in real-time as well as offline settings. In a stepping stone attack, the true origin of the attack cannot be identified directly; the only practical response is to detect the stepping stone hosts themselves. Several approaches have been proposed for this purpose. The earliest, put forward by Staniford-Chen and Heberlein (1995), was content-based: it identified stepping stones by inspecting the content of the packets passing through a host. This method, also known as the thumbprint approach, fails when the

attack traffic is encrypted, because encryption hides the packet content.

Researchers have since proposed a variety of stepping stone detection techniques (Staniford-Chen and Heberlein, 1995). Some of them decide whether a particular host is acting as a stepping stone, whereas others reconstruct the intrusion path that the attacker has built and used; detecting the path reveals all the stepping stones in one go. Despite the number of detection methods available, attackers keep finding new evasion techniques to slip past them. These include timing perturbation, where delays are introduced into packet transmission (Kesdogan et al., 1998); chaff perturbation, where dummy packets are inserted into the stream (Berthold et al., 2000; Berthold and Langos, 2002); encryption of the attack traffic so that packet content is hidden (Mishra et al., 2018; Zhang and Paxson, 2000); and reshuffling, where the order of packets is changed (Berthold et al., 2000; Pfitzmann et al., 1991).

The rest of the paper is organised as follows. Section 2 gives the background and a timeline of stepping stone attacks. Section 3 gives an overview of stepping stone attacks and the motivation behind them and their detection. Section 4 presents a taxonomy of stepping stone detection approaches, Section 5 covers performance evaluation metrics, Section 6 discusses open issues and challenges, and Section 7 concludes the paper.

2. Background and Timeline

2.1 Background

Science and technology advanced rapidly over the last century and produced new technologies in almost every field, including transportation, healthcare, aviation, and information and communication technology (ICT). ICT in particular grew very quickly during the last quarter of the century, and the Internet is among the most important technologies that scientists and researchers invented and refined in this area (Dingledine et al., 2004). Although the Internet is less than a quarter of a century old, it has already been through many challenging moments. It is under serious threat every day, and these threats affect not only its performance but also its reliability. The number of cyber-attacks carried out worldwide each year may well exceed one million and, more importantly, some attacks probably go unnoticed because the attackers use sophisticated tools and technology. To carry out an attack, an attacker exploits weaknesses in communication protocols and in computer systems. Common attack types include malware, Trojan horses, DoS and DDoS, and stepping stone attacks, among many others. The choice of attack depends on the kind of vulnerability being targeted; DoS and DDoS attacks, for example, exploit weaknesses in TCP and UDP. Table 1 lists the different vulnerabilities that attackers commonly exploit.

Table 1: Types of vulnerabilities exploited by attackers

Sr. No.	Reference	Type of Vulnerability	Description
1	Alkhalifah et al. (2019)	Software vulnerabilities	Weaknesses introduced into software applications while they are being developed. They may arise accidentally or deliberately and can change the purpose of the software altogether, turning it into a gap in the security of the system.
2	Taylor et al. (2019)	Hardware vulnerabilities	Weaknesses that stem from flaws in the design of hardware components.
3	Lee (2019)	Intersection between software and hardware component vulnerabilities	Components that sit between hardware and software, such as RAM, can be reprogrammed so that they introduce a weakness into the machine, which an attacker can then exploit for illegal activities.
4	Alkhalifah et al. (2019)	Network vulnerabilities	The connection between hosts on a network can itself be weak; such weaknesses are called connection vulnerabilities. An attacker may take over a connection and pose as its genuine user. Once the system accepts the attacker as legitimate, the connection can be used for attack activities, or it can be blocked so that no other user can use it.
5	Lee (2019)	Configuration vulnerabilities	Computers and other ICT devices can be configured in different ways, and each configuration places its own security constraints on the system components. Some configurations handle security poorly and leave the system open to attack; these are known as configuration vulnerabilities.
6	Casino et al. (2019)	User and operator vulnerabilities	Legitimate users and operators of a computer or network may be deceived or threatened into performing illegal operations.
7	Conoscenti et al. (2016)	Service provider vulnerabilities	Computer manufacturers usually obtain some components from other manufacturers. A malfunctioning component, once it becomes part of a system, can later act as a weakness through which the system is compromised.

As discussed above, stepping stone attacks pose a serious threat to the security of computers and networks. In such an attack, the attacker logs on to compromised intermediate hosts, the stepping stones, to build a connection chain and finally issues attack commands from the last system in that chain. Researchers have proposed a range of detection approaches built on the observable characteristics of

TCP/IP traffic. Correlation-coefficient and watermarking methods are commonly used to test whether an individual host is a stepping stone. Watermarking has been studied both for anonymous communication and for stepping stone detection. Houmansadr showed that the inter-packet delay within TCP/IP packet streams plays a decisive role in detecting stepping stones. Data mining methods have been applied to detect individual stepping stones as well as to stepping stone intrusion detection, which aims to recover the full length of the connection chain (the stepping stone connection chain) that the attacker has built and used. Once it became clear that detection methods relying on a single characteristic of TCP traffic were easy for attackers to defeat, researchers moved to techniques that combine several traffic characteristics. After a thorough study of the existing techniques, we selected some of the most effective approaches for stepping stone detection and stepping stone intrusion detection. They are described here together with their strengths and weaknesses, which point to research gaps for the development of new approaches.

2.2 Timeline of Stepping Stone Attacks

As mentioned earlier, stepping stones can support many kinds of cyber-attacks, including phishing, spamming, DoS, DDoS, and the spread of viruses and worms. For this reason, our account of the history of stepping stone attacks concentrates on some of the most threatening attacks carried out through stepping stones. DDoS attacks are among the most destructive and frequent of these, and it is hard to imagine a DDoS attack being carried out without stepping stones. A DDoS attack relies on many machines to generate attack traffic: the attacker first compromises a few machines, uses them to compromise others, and finally directs all the compromised machines to carry out the attack.

We therefore begin with DDoS attacks, the most destructive and most common cyber-attacks performed through stepping stones.

Over the course of this study we found that the history of stepping stone attacks includes a very large number of incidents, far too many to cover here. We have therefore listed only a few of the best-known and most destructive attacks (Table 2).

Table 2: Notable cyber-attacks carried out using stepping stones

Attack Description
The first DDoS attack was launched by an Italian group of attackers against the nuclear policy of the French government .
Another major DDoS attack was carried out in 1999 against the University of Minnesota. One of the largest cyber-attacks in history was the DDoS attack on Spamhaus, a non-profit anti-spam organisation, in 2013.
The attack slowed down the Internet across the globe and peaked at 300 Gbps. Named Stophaus, it continued from 18 March to 27 March 2013.
Another phishing attack that relied on compromised intermediate hosts (stepping stones) targeted *PayPal*, the largest online payment and money-transfer company in the world.

In this attack, the attackers used the highly secure website of the Malaysian police to host a phishing *PayPal* website. The *SSL* certificate involved could not be revoked in most browsers, including Firefox and Safari, and when a certificate cannot be revoked, the certification authority (CA) can no longer guarantee secure communication or control whether browsers accept the certificate.
Viruses and worms have also been spread through stepping stones. Stuxnet, a cyber worm built with sophisticated engineering over a long period, was used to sabotage the Iranian nuclear programme in 2009-10. It was designed to disrupt nuclear centrifuges running on a SCADA system developed by Siemens. In this attack, the Iranian website "Natanz" served as an operating platform for Stuxnet.

3. Overview of Stepping Stone Attacks and Motivation

Theft is an old idea, and as times have changed and new technologies have appeared, thieves have changed the way they steal, always looking for a safe route to follow. In the early days of computing, theft was hardly a concern in this field. As newer generations of computers arrived and computers became easier to use and more widely available, people began to store useful and confidential information on them. Before the Internet, a thief needed physical access to a computer, and its password, to steal the information stored on it. Internet thieves, usually called hackers or attackers, look for weaknesses in the technologies used in computers and networks and exploit them to carry out offensive cyber operations. Because the Internet is everywhere, attackers can reach any connected computer remotely and can therefore get at the information stored on it with relative ease. Proxies were first designed so that users could access the Internet without revealing their identity, but attackers now use them to stay anonymous while carrying out attacks. Advances in anonymity technologies help attackers hide their identity during attacks, and as these technologies develop, cyber-attacks are increasing at a faster pace. Every year a large number of attacks succeed; some cause major damage to their victims, while others are minor. In most cyber-attacks, the attacker issues commands through intermediate computers to remain anonymous, using tools such as Telnet, OpenSSH and rlogin. These remote login tools allow the attacker to log in to a networked computer, take control of it, use it to log in to another computer and take control of that one as well, and so on. In this way the attacker builds a chain of compromised systems and launches attack commands from one of them. Every host in this chain that the attacker logs into from a previously compromised host, together with the first host reached directly from the attacker's own computer, is called a stepping stone host.

Consider, for example, an attacker who builds a stepping stone connection chain made up of three stepping stone hosts, namely Host1, Host2 and Host3. The attacker first logs in remotely to Host1 using a remote login tool such as Telnet, OpenSSH or rlogin, compromises it and takes over control. With Host1 fully under control, the attacker uses it to compromise Host2, and then uses Host2 to compromise Host3. The attack commands are finally launched from Host3. During an investigation, Host3 will appear to be the origin of the attack: owing to the way TCP/IP works, the attack packets that reach the victim carry the IP address of the last host in the chain, so Host3 is taken to be the

attacker.

The growth of online shopping and net banking has encouraged attackers to carry out fraudulent financial transactions. They frequently target payment gateways to obtain transaction data, which may contain user IDs and passwords that can then be used to enter user accounts and make transactions. Stepping stone hosts are generally used for these activities so that the attacker is shielded from detection. Attackers also use stepping stones to obtain services and resources without paying, for instance by buying goods on shopping sites without paying for them. Most developed and developing countries now keep their information in digital form, accessible through secure communication channels, and an adversary may try to reach the confidential information of another country through stepping stone hosts located in a third country. All of this shows how serious stepping stone attacks are. It falls to researchers to develop effective real-time stepping stone detection techniques that can counter these attacks, protect networks and computers, and keep the machines of innocent users from being turned into stepping stones.

4. Taxonomy of Stepping Stone Detection Approaches

Stepping stone detection approaches can be grouped in several ways. Based on the information they examine, they are either content-based, relying on the payload of packets, or timing-based, relying on packet timing, packet counts and other traffic characteristics. Based on how they observe traffic, they are either passive, monitoring traffic as it is, or active, deliberately modifying traffic so that it can be recognised further downstream. Based on their goal, they perform either stepping stone detection (SSD), which decides whether a host is relaying an interactive connection, or stepping stone intrusion detection (SSID), which estimates the length of the downstream connection chain to decide whether a session is part of an intrusion. Approaches also differ in where measurements are taken: some need monitors at several points in the network, while others work from a single host. Earlier surveys (Shullich et al., 2011; Wang and Yang, 2018) give broader coverage; this section focuses on the main families of approaches and their trade-offs, which are summarised in Table 3.

4.1 Content-Based Approaches

The earliest approach to stepping stone detection was proposed by Staniford-Chen and Heberlein (1995). It computes thumbprints, which are short summaries of the content of a connection over a given time interval, and compares the thumbprints of different connections. Connections whose thumbprints closely match are assumed to carry the same data and therefore to belong to the same chain. Thumbprints are compact and cheap to compare, which makes them suitable for monitoring a large number of connections. However, the approach assumes that packet content stays the same as it passes through each host. Once an attacker uses an encrypted remote login tool such as SSH, the content on each hop is encrypted with a different key, the thumbprints no longer match, and the method fails. Because encryption is now the norm for remote access, content-based methods are mainly of historical interest.

4.2 Timing-Based Passive Approaches

To cope with encrypted traffic, later approaches use

only information that encryption does not hide, chiefly the timing of packets. Zhang and Paxson (2000) observed that interactive sessions alternate between ON periods, when the user is typing, and OFF periods, when the connection is idle. If two connections on a host are part of the same chain, the end of an OFF period on one is followed closely by the end of an OFF period on the other. By counting such coincidences, the method flags pairs of connections that are likely to be relayed, with a low false positive rate and modest processing cost. Yoda and Etoh (2000) took a related route, computing the deviation between the packet streams of two connections and treating a small deviation as evidence that they belong to the same chain.

Wang et al. (2002) proposed correlating the inter-packet delays (IPDs) of connections. Because a relaying host forwards each keystroke soon after it arrives, the IPD sequences of related connections are strongly correlated, and the authors showed that a few hundred packets are often enough for a reliable decision. These passive methods are, however, vulnerable to deliberate evasion. An attacker can add random delays to packets (timing perturbation) or insert dummy packets (chaff) that are discarded at the next hop, which disturbs the timing patterns on which the methods depend.

Several studies have addressed these evasion techniques directly. Donoho et al. (2002) used multiscale wavelet analysis and showed that, if the attacker can delay packets only up to a maximum tolerable delay, the correlation between related connections cannot be hidden at sufficiently long time scales. Blum et al. (2004) proposed packet-counting algorithms with provable bounds on the false positive rate and on the number of packets that must be observed, and extended them to tolerate a bounded amount of chaff. He and Tong (2007) developed detectors for encrypted connections under bounded delay and chaff and analysed how much chaff an attacker needs to evade detection. A common conclusion of this work is that robustness comes at the cost of longer observation periods. More recently, Clausen et al. (2020) released a large dataset of simulated stepping stone traffic and used it to evaluate eight state-of-the-art methods; several fell short of their reported detection rates, and none could reliably detect stepping stones once the attacker added suitable chaff.

4.3 Active Watermark-Based Approaches

Active approaches do not wait for natural patterns to appear; they create them. Wang and Reeves (2003) proposed embedding a watermark in the IPDs of a flow by slightly delaying selected packets as they leave a monitored point. If the same watermark is found in a flow observed elsewhere, the two flows are linked. Because the watermark is designed to survive random delays, this approach is more robust to timing perturbation than passive correlation and needs fewer packets to reach a decision. Later schemes such as RAINBOW (Houmansadr et al., 2009) reduce the delays needed, so that the watermark is harder for the attacker to notice. Active approaches have their own drawbacks. They require control over traffic at the point where the watermark is embedded, they add latency to legitimate traffic, and an attacker who can detect or remove the watermark can defeat them (Peng et al., 2006). Chaff and repacketisation can also weaken the embedded pattern.

4.4 Connection-Chain Length Estimation

Methods that only decide whether a host is a stepping stone cannot tell an attacker from a legitimate user who simply logs in through one or two intermediate hosts, which is common practice among system administrators. SSID methods address this by estimating the length of the connection chain downstream of the monitored host, on the assumption that intruders use long chains while legitimate users rarely go beyond a few hops. Yung (2002) compared the delay of the acknowledgement of a keystroke with the delay of its echo to estimate whether a session passes through a long chain. Yang and Huang (2007) matched send and echo packets to obtain round-trip times (RTTs) and clustered them to estimate the number of hosts in the chain, and Ding et al. (2009) likewise targeted intruders who use long connection chains. Recent work has refined the packet-matching and RTT-analysis steps, for example through k-means clustering of RTTs (Wang et al., 2021), MMD data mining to match send and echo packets (Yang and Wang, 2021) and analysis of the RTT distribution of network traffic (Yang and Wang, 2024). These approaches have the practical advantage of working from a single monitoring point. Their accuracy, however, depends on matching send and echo packets correctly, which is difficult when packets are encrypted, merged or split, and on RTT measurements that fluctuate with network load.

4.5 Data Mining and Machine Learning Approaches

Data mining and machine learning have been used both to detect stepping stones and to estimate chain length. Hsiao and Fan (2009) applied network traffic mining to identify stepping stone connections, while Wu and Huang (2010) trained neural networks on packet timing features to detect stepping stone intrusion. Learning-based methods can combine several traffic characteristics at once, such as timing, packet size and RTT, which makes them harder to evade than methods that rely on a single feature. Deep learning flow-correlation models, first developed to link flows in the Tor network (Nasr et al., 2018), have recently been applied to stepping stone intrusion detection: Mathews et al. (2026) report a true positive rate above 0.99 at a false positive rate of 10^{-3} across several tunnelling protocols, although their evaluation relies on synthetic traffic. The performance of learning-based methods, however, depends on the quality and representativeness of the training data, and a model trained in one network environment may not generalise to another. They are also open to adversarial manipulation, in which the attacker shapes traffic specifically to mislead the model.

4.6 Relevance to Blockchain Networks

The approaches above were developed for interactive sessions over TCP/IP, but they apply directly to the infrastructure on which blockchain systems run. Full nodes, mining pool servers, wallet services, exchanges and remote procedure call (RPC) endpoints are all networked hosts that can be compromised and then used as stepping stones, either to attack other participants or to reach high-value targets such as exchange hot wallets. Attackers who reach blockchain infrastructure through anonymity networks such as Tor (Dingledine et al., 2004) or through chains of compromised hosts are difficult to trace, and the pseudonymous nature of blockchain addresses adds to

this difficulty. Network-level stepping stone detection therefore complements the ledger-level analysis commonly used to investigate blockchain attacks. Some characteristics of blockchain networks, discussed in Section 6, also make detection harder.

Table 3: Summary of stepping stone detection approaches

Approach	**Representative work**	**Main limitation**
Content-based (thumbprinting)	Staniford-Chen and Heberlein (1995)	Fails on encrypted traffic
ON/OFF timing correlation	Zhang and Paxson (2000)	Sensitive to timing perturbation and chaff
Deviation-based	Yoda and Etoh (2000)	Sensitive to delays; needs long observation
IPD correlation	Wang et al. (2002)	Degrades under jitter and chaff
Multiscale analysis and packet counting	Donoho et al. (2002); Blum et al. (2004); He and Tong (2007)	Many packets needed for a confident decision
Active watermarking	Wang and Reeves (2003); Houmansadr et al. (2009)	Requires traffic manipulation; watermark may be detected or removed
Chain-length (RTT) estimation	Yung (2002); Yang and Huang (2007); Ding et al. (2009); Yang and Wang (2024)	Noisy RTTs; send-echo matching is difficult
Data mining and machine learning	Hsiao and Fan (2009); Wu and Huang (2010); Mathews et al. (2026)	Needs labelled data; open to adversarial evasion

5. Performance Evaluation Metrics

Most stepping stone detection problems can be framed as binary classification: a pair of connections either is or is not part of the same chain, and a host either is or is not a stepping stone. Evaluation therefore draws on the metrics used for intrusion detection systems in general, together with metrics that capture robustness to evasion and the cost of detection. Table 4 summarises the main metrics.

5.1 Detection Accuracy

Detection accuracy is measured from the counts of true positives (TP), false positives (FP), true negatives (TN) and false negatives (FN). The true positive rate (TPR), also called the detection rate or recall, is the fraction of actual stepping stone connections that are detected. The false positive rate (FPR) is the fraction of unrelated connection pairs that are wrongly flagged. Precision and the F1-score summarise the trade-off between the two, and receiver operating characteristic (ROC) curves show how the TPR and FPR change as the detection threshold varies. The FPR deserves particular attention. On a busy host the number of unrelated connection pairs is far larger than the number of relayed ones, so even a small

FPR can produce more false alarms than true detections. Overall accuracy is therefore a misleading measure in this setting and should be reported alongside the FPR and precision.

5.2 Robustness to Evasion

Because attackers actively try to evade detection, a useful evaluation reports how detection performance changes under evasion. Typical measures are the detection rate as a function of the maximum delay the attacker adds to packets, the ratio of chaff packets to genuine packets, and the presence of encryption, repacketisation and packet loss. Reporting results across a range of perturbation levels shows where a method breaks down, which is more informative than a single figure obtained under ideal conditions.

5.3 Efficiency and Practicality

A detector must also be practical to run. Relevant measures include the number of packets, or the observation time, needed before a reliable decision can be made; the computational cost, which grows quickly when every pair of connections on a host must be compared; memory requirements; and whether the method can operate in real time. For active approaches, the delay added to legitimate traffic and the visibility of the watermark to an attacker are further costs. For SSID methods, the accuracy of the estimated chain length, for example the mean absolute error between the estimated and actual number of hops, is the key measure.

Table 4: Performance evaluation metrics for stepping stone detection

Metric	Definition
True positive rate (detection rate, recall)	TP / (TP + FN)
False positive rate	FP / (FP + TN)
Precision	TP / (TP + FP)
F1-score	2 × Precision × TPR / (Precision + TPR)
Accuracy	(TP + TN) / (TP + TN + FP + FN)
ROC curve and AUC	TPR plotted against FPR as the threshold varies; AUC is the area under this curve
Robustness	Detection rate under a given maximum delay, chaff ratio or encryption setting
Packets to decision	Number of packets (or observation time) needed for a reliable decision
Chain-length error	Difference between the estimated and actual number of hops

6. Open Issues and Challenges

Despite more than two decades of research, stepping stone detection is far from a solved problem. The main open issues are outlined below.

6.1 Evasion through Perturbation and Encryption

Encryption has made content-based detection obsolete, and timing perturbation and chaff continue to challenge timing-based methods. Theoretical results show that detection remains possible when the attacker's delay and chaff are bounded (Donoho et al., 2002; Blum et al., 2004), but the number of packets needed grows as the perturbation increases, and an attacker who combines delay, chaff and repacketisation may stay below the detection threshold for the duration of a short attack. Methods that remain reliable under combined evasion techniques, with short observation times, are still needed.

6.2 Real-Time Detection and Scalability

Many methods compare every pair of incoming and outgoing connections on a host, and the cost of doing so grows rapidly with traffic volume. On high-speed links and busy servers, which are exactly the kinds of hosts attackers prefer, this cost limits real-time use. Efficient filtering of candidate pairs, streaming algorithms and hardware acceleration, including detection inside programmable network switches (Bhattacherjee et al., 2019), are possible directions.

6.3 Distinguishing Intrusions from Legitimate Relaying

Legitimate users and administrators often reach systems through jump hosts and bastion servers, and such relaying looks much like a stepping stone chain. SSID methods that estimate chain length address part of this problem, but RTT-based estimates are noisy and attackers can choose to use short chains. Combining chain length with other evidence, such as host behaviour and access policies, could reduce false alarms.

6.4 Multiple Monitoring Points and Cross-Domain Cooperation

Tracing a complete chain requires observations from many points, yet the hosts of a chain may lie in different organisations and countries (Hsiao and Fan, 2009). Sharing traffic information across administrative and national boundaries raises technical, legal and privacy problems, and there is no widely adopted framework for cooperative traceback.

6.5 Datasets and Reproducible Evaluation

There are few public datasets of stepping stone traffic, and most studies evaluate their methods on their own testbeds or on simulated chains built under differing assumptions. This makes fair comparison between methods difficult. Clausen et al. (2020) released a large dataset for this purpose, and Mathews et al. (2026) developed a tool that generates synthetic stepping stone traffic over several protocols, but more shared resources of this kind are needed. Shared benchmark datasets that include encrypted traffic, realistic background traffic and a range of evasion techniques, together with common evaluation metrics such as those described in Section 5, would help the field progress.

6.6 Robustness of Learning-Based Detectors

Machine learning methods can combine many traffic features, but they need labelled training data, may not transfer from one network to another, and can be targeted by adversarial traffic shaping. Their decisions are also often hard to explain, which matters when detection results are used as evidence in investigations.

6.7 Challenges Specific to Blockchain Networks

Blockchain networks raise further issues. Peer-to-peer nodes routinely relay transactions and blocks to their neighbours, so relaying is part of normal operation and may be mistaken for stepping stone activity. Many nodes deliberately run behind anonymity networks such as Tor to protect the privacy of their operators, which also benefits attackers. Because blockchain networks are decentralised, there is no single administrator who can deploy monitors across the network, and many participants are known only by pseudonymous addresses. At the same time, blockchain technology offers an opportunity: tamper-resistant ledgers could be used to record and share traceback evidence among cooperating organisations without relying on a single trusted party. Detection methods adapted to the traffic patterns of blockchain protocols remain largely unexplored.

6.8 Privacy and Ethical Concerns

Detection requires monitoring user traffic, and active methods go further by modifying it. Approaches that protect the privacy of legitimate users, for example by analysing only timing metadata or by processing data locally, need to be balanced against the need for effective detection.

7. Conclusion

Stepping stones allow attackers to launch attacks through chains of compromised hosts while hiding their true origin, and they have been used in DDoS attacks, phishing campaigns and the spread of worms. As blockchain systems, exchanges and related services become attractive targets, the same techniques threaten the infrastructure on which these systems depend. This paper has reviewed the background and history of stepping stone attacks and presented a taxonomy of detection approaches covering content-based, passive timing-based, active watermark-based, chain-length estimation, and data mining and machine learning methods. It has also described the metrics used to evaluate these approaches and the main open issues.

The review shows a clear progression. Content-based methods gave way to timing-based methods once encryption became common, and passive timing methods were followed by more robust statistical, active and learning-based methods as attackers adopted timing perturbation and chaff. No single approach is yet robust to combined evasion techniques while remaining fast, scalable and accurate. Future work should focus on methods that resist combined evasion with short observation times, on shared benchmark datasets, on cooperative traceback across domains, and on detection methods suited to decentralised blockchain networks. We hope this survey will serve as a useful starting point for researchers working towards these goals.

References

Alkhalifah A, et al. 2019. A taxonomy of blockchain threats and vulnerabilities.

Arachchilage NAG, Love S. 2013. A game design framework for avoiding phishing attacks. Comput Hum Behav. 29(3):706-714.

Arachchilage NAG, Love S. 2014. Security awareness of computer users: a phishing threat avoidance

perspective. Comput Hum Behav. 38:304-312.

Berthold O, Federrath H, Kopsell S. 2000. Web MIXes: a system for anonymous and unobservable Internet access. In: Proceedings of the Workshop on Design Issues in Anonymity and Unobservability. Lecture Notes in Computer Science, vol. 2009. Berkeley, CA, pp. 115-129.

Berthold O, Langos H. 2002. Dummy traffic against long term intersection attacks. In: Proceedings of the Privacy Enhancing Technologies Workshop (PET'02). Lecture Notes in Computer Science, vol. 2482. San Francisco, CA, pp. 110-128.

Bhattacherjee D, Gurtov A, Aura T. 2019. Watch your step! Detecting stepping stones in programmable networks. In: Proceedings of the IEEE International Conference on Communications (ICC 2019). IEEE, pp. 1-7.

Blum A, Song D, Venkataraman S. 2004. Detection of interactive stepping stones: algorithms and confidence bounds. In: Recent Advances in Intrusion Detection (RAID 2004). Lecture Notes in Computer Science, vol. 3224. Springer, pp. 258-277.

Casino F, Dasaklis TK, Patsakis C. 2019. A systematic literature review of blockchain-based applications: current status, classification and open issues. Telemat Inform. 36:55-81.

Chaum DL. 1981. Untraceable electronic mail, return addresses, and digital pseudonyms. Commun ACM. 422-426.

Clausen H, Gibson MS, Aspinall D. 2020. Evading stepping-stone detection with enough chaff. In: Network and System Security (NSS 2020). Lecture Notes in Computer Science, vol. 12570. Springer, pp. 431-446.

Conoscenti M, Vetro A, De Martin JC. 2016. Blockchain for the Internet of Things: a systematic literature review. In: 2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA). IEEE.

Ding W, et al. 2009. Detecting stepping-stone intruders with long connection chains. In: Proceedings of the 5th IEEE International Conference on Information Assurance and Security. IEEE, pp. 665-669.

Dingledine R, Mathewson N, Syverson P. 2004. Tor: the second-generation onion router. In: Proceedings of the 13th USENIX Security Symposium. San Diego, CA, USA, pp. 21-38.

Donoho DL, Flesia AG, Shankar U, Paxson V, Coit J, Staniford S. 2002. Multiscale stepping-stone detection: detecting pairs of jittered interactive streams by exploiting maximum tolerable delay. In: Recent Advances in Intrusion Detection (RAID 2002). Lecture Notes in Computer Science, vol. 2516. Springer, pp. 17-35.

Gupta S, Gupta BB. 2015. XSS-SAFE: a server-side approach to detect and mitigate cross-site scripting (XSS) attacks in JavaScript code. Arab J Sci Eng. 1-24.

Gupta S, Gupta BB. 2016. JS-SAN: defense mechanism for HTML5-based web applications against JavaScript code injection vulnerabilities. Secur Commun Netw. 26.

He T, Tong L. 2007. Detecting encrypted stepping-stone connections. IEEE Trans Signal Process. 55(5):1612-1623.

Houmansadr A, Kiyavash N, Borisov N. 2009. RAINBOW:

a robust and invisible non-blind watermark for network flows. In: Proceedings of the 16th Network and Distributed System Security Symposium (NDSS). San Diego, CA.

Hsiao H, Fan WC. 2009. Detecting step stone with network traffic mining approach. In: Proceedings of the 4th IEEE Conference on Innovative Computing, Information and Control (ICICIC). IEEE, pp. 1176-1179.

Internet Society. 2020. A brief history of the Internet & related networks. Available online: http://www.internetsociety.org/internet/what-internet/history-internet/brief-history-internet-related-networks [Accessed 30 April 2020].

Kesdogan D, Egner J, Buschkes R. 1998. Stop-and-go MIXes providing probabilistic security in an open system. In: Second International Workshop on Information Hiding (IH'98). Lecture Notes in Computer Science, vol. 1525. Portland, OR, pp. 83-98.

Kushner D. 2013. The real story of Stuxnet. IEEE Spectrum. Available online: http://spectrum.ieee.org/telecom/security/the-real-story-of-stuxnet [Accessed November 2014].

Lai YP, Hsia PL. 2007. Using the vulnerability information of computer systems to improve the network security. Comput Commun. 30(9):2032-2047.

Lee JH. 2019. Systematic approach to analyzing security and vulnerabilities of blockchain systems. Dissertation. Massachusetts Institute of Technology.

Mason. n.d. Hackers are using Malaysian police website for PayPal phishing attacks. Available online: https://traxarmstrong.com/malaysian-police-site-is-been-used-for-paypal-phishing/ [Accessed March 2015].

Mathews N, Hopper N, Wright M. 2026. Tracing the chain: deep learning for stepping-stone intrusion detection. arXiv preprint arXiv:2604.08800.

Mishra A, Gupta BB, Gupta D. 2018. Identity theft, malware, and social engineering in dealing with cybercrime. In: Computer and Cyber Security. Auerbach Publications, pp. 627-648.

Nasr M, Bahramali A, Houmansadr A. 2018. DeepCorr: strong flow correlation attacks on Tor using deep learning. In: Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security (CCS). Toronto, Canada, pp. 1962-1976.

Peng P, Ning P, Reeves DS. 2006. On the secrecy of timing-based active watermarking trace-back techniques. In: Proceedings of the IEEE Symposium on Security and Privacy. Berkeley, CA, pp. 334-349.

Pfitzmann A, Pfitzmann B, Waidner M. 1991. ISDN-MIXes: untraceable communication with very small bandwidth overhead. In: Proceedings of the GI/ITG Conference: Communication in Distributed Systems. Informatik-Fachberichte, vol. 267. Mannheim, Germany, pp. 451-463.

Popescu R. n.d. Facebook apps hosted by Heroku used for viral Twitter phishing attack. Netcraft. Available online: http://news.netcraft.com/page/18/ [Accessed March 2015].

Sauter M. n.d. The history of DDoS attacks as a tool of protest. Motherboard. Available online: http://motherboard.vice.com/read/history-of-the-ddos-attack [Accessed March 2015].

Shahmehri N, et al. 2012. An advanced approach for modeling and detecting software vulnerabilities. Inf Softw Technol. 54(9):997-1013.

Shullich R, Chu J, Ji P, Chen W. 2011. A survey of research in stepping-stone detection. Int J Electron Commer Stud. 2(2):103-126.

Staniford-Chen S, Heberlein LT. 1995. Holding intruders accountable on the Internet. In: Proceedings of the IEEE Symposium on Security and Privacy. Oakland, CA, pp. 39-49.

Taylor PJ, et al. 2019. A systematic literature review of blockchain cyber security. Digit Commun Netw.

Wang L, Yang J. 2018. A research survey in stepping-stone intrusion detection. EURASIP J Wirel Commun Netw. 2018:276.

Wang L, Yang J, Xu X, Wan PJ. 2021. Mining network traffic with the k-means clustering algorithm for stepping-stone intrusion detection. Wirel Commun Mob Comput. 2021:6632671.

Wang X, Reeves DS. 2003. Robust correlation of encrypted attack traffic through stepping stones by manipulation of interpacket delays. In: Proceedings of the 10th ACM Conference on Computer and Communications Security (CCS). Washington, DC, pp. 20-29.

Wang X, Reeves DS, Wu SF. 2002. Inter-packet delay based correlation for tracing encrypted connections through stepping stones. In: Proceedings of the 7th European Symposium on Research in Computer Security (ESORICS 2002). Lecture Notes in Computer Science, vol. 2502. Zurich, Switzerland, pp. 244-263.

Wu HC, Huang SHS. 2010. Neural networks-based detection of stepping-stone intrusion. Expert Syst Appl. 37(2):1431-1437.

Yang J, Huang SHS. 2007. Mining TCP/IP packets to detect stepping-stone intrusion. Comput Secur. 26(7-8):479-484.

Yang J, Wang L. 2021. Applying MMD data mining to match network traffic for stepping-stone intrusion detection. Sensors. 21(22):7464.

Yang J, Wang L. 2024. Explore utilizing network traffic distribution to detect stepping-stone intrusion. Electronics. 13(16):3258.

Yoda K, Etoh H. 2000. Finding a connection chain for tracing intruders. In: Proceedings of the 6th European Symposium on Research in Computer Security (ESORICS 2000). Lecture Notes in Computer Science, vol. 1895. Toulouse, France, pp. 191-205.

Yung KH. 2002. Detecting long connection chains of interactive terminal sessions. In: Recent Advances in Intrusion Detection (RAID 2002). Lecture Notes in Computer Science, vol. 2516. Springer, pp. 1-16.

Zhang Y, Paxson V. 2000. Detecting stepping-stones. In: Proceedings of the 9th USENIX Security Symposium. Denver, CO, pp. 67-8