

Internet Draft
draft-ietf-tcpm-fastopen-04.txt
Intended status: Experimental
Expiration date: February, 2014

Y. Cheng
J. Chu
S. Radhakrishnan
A. Jain
Google, Inc.
July 15, 2013

TCP Fast Open

Status of this Memo

Distribution of this memo is unlimited.

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/lid-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

TCP Fast Open (TFO) allows data to be carried in the SYN and SYN-ACK packets and consumed by the receiving end during the initial connection handshake, thus saving up to one full round trip time (RTT) compared to the standard TCP, which requires a three-way handshake (3WHS) to complete before data can be exchanged. However TFO deviates from the standard TCP semantics in that the data in the SYN could be replayed to an application in some rare circumstances. Applications should not use TFO unless they can tolerate this issue, which is detailed in the Applicability section.

Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119]. TFO refers to TCP Fast Open. Client refers to the TCP's active open side and server refers to the TCP's passive open side.

Table of Contents

1. Introduction	3
2. Data In SYN	3
2.1 Relaxing TCP Semantics on Duplicated SYNs	4
2.2. SYNs with Spoofed IP Addresses	4
3. Protocol Overview	5
4. Protocol Details	7
4.1. Fast Open Cookie	7
4.1.1. TCP Options	7
4.1.2. Server Cookie Handling	8
4.1.3. Client Cookie Handling	9
4.2. Fast Open Protocol	9
4.2.1. Fast Open Cookie Request	10
4.2.2. TCP Fast Open	11
5. Security Considerations	13
5.1. Resource Exhaustion Attack by SYN Flood with Valid Cookies	13
5.1.1 Attacks from behind Sharing Public IPs (NATs)	14
5.2. Amplified Reflection Attack to Random Host	15
6. TFO's Applicability	16
6.1 Duplicate Data in SYNs	16
6.2 Potential Performance Improvement	16
6.3. Example: Web Clients and Servers	16
6.3.1. HTTP Request Replay	16
6.3.2. Comparison with HTTP Persistent Connections	17
7. Open Areas for Experimentation	17
7.1. Performance impact due to middle-boxes and NAT	18
7.2. Cookie-less Fast Open	18
8. Related Work	18

8.1. T/TCP	19
8.2. Common Defenses Against SYN Flood Attacks	19
8.3. TCP Cookie Transaction (TCPCT)	19
8.4. Speculative Connections by the Applications	19
9. IANA Considerations	19
10. Acknowledgement	20
11. References	20
11.1. Normative References	20
11.2. Informative References	20
Appendix A. Example Socket API Changes to support TFO	22
Authors' Addresses	22

1. Introduction

TCP Fast Open (TFO) enables data to be exchanged safely during TCP's connection handshake. This document describes a design that enables applications to save a round trip while avoiding severe security ramifications. At the core of TFO is a security cookie used by the server side to authenticate a client initiating a TFO connection. This document covers the details of exchanging data during TCP's initial handshake, the protocol for TFO cookies, potential new security vulnerabilities and their mitigation, and the new socket API.

TFO is motivated by the performance needs of today's Web applications. Current TCP only permits data exchange after 3WSH [RFC793], which adds one RTT to network latency. For short Web transfers this additional RTT is a significant portion of overall network latency [THK98], even when HTTP persistent connection is widely used. For example, the Chrome browser keeps TCP connections idle for up to 5 minutes but 35% of Chrome HTTP requests are made on new TCP connections [RCCJR11]. For such Web and Web-like applications placing data in the SYN can yield significant latency improvements. Next we describe how we resolve the challenges that arise upon doing so.

2. Data In SYN

Standard TCP already allows data to be carried in SYN packets ([RFC793], section 3.4) but forbids the receiver from delivering it to the application until 3WSH is completed. This is because TCP's initial handshake serves to capture old or duplicate SYNs.

Allowing data in SYN packets to be delivered raises two issues discussed in the following subsections. These issues make TFO unsuitable for certain applications. Therefore TCP implementations

MUST NOT use TFO by default, but only use TFO if requested explicitly by the application on a per service port basis. Applications need to evaluate TFO applicability described in Section 6 before using TFO.

2.1 Relaxing TCP Semantics on Duplicated SYNs

TFO allows data to be delivered to the application before 3WSH is completed, thus opening itself to a data integrity issue in either of the two cases below:

a) the receiver host receives data in a duplicate SYN after it has forgotten it received the original SYN (e.g. due to a reboot);

b) the duplicate is received after the connection created by the original SYN has been closed and the close was initiated by the sender (so the receiver will not be protected by the 2MSL TIMEWAIT state).

The now obsoleted T/TCP [RFC1644] attempted to address these issues. It is not successful and not deployed due to various vulnerabilities [PHRACK98]. Rather than trying to capture all dubious SYN packets to make TFO 100% compatible with TCP semantics, we made a design decision early on to accept old SYN packets with data, i.e., to restrict TFO use to a class of applications (Section 6) that are tolerant of duplicate SYN packets with data. We believe this is the right design trade-off balancing complexity with usefulness.

2.2. SYNs with Spoofed IP Addresses

Standard TCP suffers from the SYN flood attack [RFC4987] because bogus SYN packets, i.e., SYN packets with spoofed source IP addresses can easily fill up a listener's small queue, causing a service port to be blocked completely until timeouts. Secondary damage comes from these SYN requests taking up memory space. Though this is less of an issue today as servers typically have plenty of memory.

TFO goes one step further to allow server-side TCP to send up data to the application layer before 3WSH is completed. This opens up serious new vulnerabilities. Applications serving ports that have TFO enabled may waste lots of CPU and memory resources processing the requests and producing the responses. If the response is much larger than the request, the attacker can further mount an amplified reflection attack against victims of choice beyond the TFO server itself.

Numerous mitigation techniques against regular SYN flood attacks exist and have been well documented [RFC4987]. Unfortunately none are applicable to TFO. We propose a server-supplied cookie to mitigate these new vulnerabilities in Section 3 and evaluate the effectiveness

of the defense in Section 7.

3. Protocol Overview

The key component of TFO is the Fast Open Cookie (cookie), a message authentication code (MAC) tag generated by the server. The client requests a cookie in one regular TCP connection, then uses it for future TCP connections to exchange data during 3WHS:

Requesting a Fast Open Cookie:

1. The client sends a SYN with a Fast Open Cookie Request option.
2. The server generates a cookie and sends it through the Fast Open Cookie option of a SYN-ACK packet.
3. The client caches the cookie for future TCP Fast Open connections (see below).

Performing TCP Fast Open:

1. The client sends a SYN with Fast Open Cookie option and data.
2. The server validates the cookie:
 - a. If the cookie is valid, the server sends a SYN-ACK acknowledging both the SYN and the data. The server then delivers the data to the application.
 - b. Otherwise, the server drops the data and sends a SYN-ACK acknowledging only the SYN sequence number.
3. If the server accepts the data in the SYN packet, it may send the response data before the handshake finishes. The max amount is governed by the TCP's congestion control [RFC5681].
4. The client sends an ACK acknowledging the SYN and the server data. If the client's data is not acknowledged, the client retransmits the data in the ACK packet.
5. The rest of the connection proceeds like a normal TCP connection. The client can repeat many Fast Open operations once it acquires a cookie (until the cookie is expired by the server). Thus TFO is useful for applications that have temporal locality on client and server connections.

Requesting Fast Open Cookie in connection 1:

TCP A (Client)		TCP B(Server)
CLOSED		LISTEN
#1 SYN-SENT	----- <SYN, CookieOpt=NIL> ----->	SYN-RCVD
#2 ESTABLISHED (caches cookie C)	<----- <SYN, ACK, CookieOpt=C> -----	SYN-RCVD

Performing TCP Fast Open in connection 2:

TCP A (Client)		TCP B(Server)
CLOSED		LISTEN
#1 SYN-SENT	----- <SYN=x, CookieOpt=C, DATA_A> ----->	SYN-RCVD
#2 ESTABLISHED	<----- <SYN=y, ACK=x+len(DATA_A)+1> -----	SYN-RCVD
#3 ESTABLISHED	<----- <ACK=x+len(DATA_A)+1, DATA_B>-----	SYN-RCVD
#4 ESTABLISHED	----- <ACK=y+1>----->	ESTABLISHED
#5 ESTABLISHED	--- <ACK=y+len(DATA_B)+1>----->	ESTABLISHED

from the Fast Open cookie option.
Options with invalid Length values, without SYN flag set, or with ACK flag set MUST be ignored.

4.1.2. Server Cookie Handling

The server is in charge of cookie generation and authentication. The cookie SHOULD be a message authentication code tag with the following properties:

1. The cookie authenticates the client's (source) IP address of the SYN packet. The IP address may be an IPv4 or IPv6 address.
2. The cookie can only be generated by the server and can not be fabricated by any other parties including the client.
3. The generation and verification are fast relative to the rest of SYN and SYN-ACK processing.
4. A server may encode other information in the cookie, and accept more than one valid cookie per client at any given time. But this is all server implementation dependent and transparent to the client.
5. The cookie expires after a certain amount of time. The reason for cookie expiration is detailed in the "Security Consideration" section. This can be done by either periodically changing the server key used to generate cookies or including a timestamp when generating the cookie.

To gradually invalidate cookies over time, the server can implement key rotation to generate and verify cookies using multiple keys. This approach is useful for large-scale servers to retain Fast Open rolling key updates. We do not specify a particular mechanism because the implementation is server specific.

The server supports the cookie generation and verification operations:

- GetCookie(IP_Address): returns a (new) cookie
- IsCookieValid(IP_Address, Cookie): checks if the cookie is valid, i.e., it has not expired and it authenticates the client IP address.

Example Implementation: a simple implementation is to use AES_128 to encrypt the IPv4 (with padding) or IPv6 address and truncate to 64 bits. The server can periodically update the key to expire the cookies. AES encryption on recent processors is fast and takes only a

few hundred nanoseconds [RCCJR11].

If only one valid cookie is allowed per-IP and the server can regenerate the cookie independently, the best validation process is to simply regenerate a valid cookie and compare it against the incoming cookie. In that case if the incoming cookie fails the check, a valid cookie is readily available to be sent to the client.

4.1.3. Client Cookie Handling

The client **MUST** cache cookies from servers for later Fast Open connections. For a multi-homed client, the cookies are both client and server IP dependent. Beside the cookie, we **RECOMMEND** that the client caches the MSS and RTT to the server to enhance performance.

The MSS advertised by the server is stored in the cache to determine the maximum amount of data that can be supported in the SYN packet. This information is needed because data is sent before the server announces its MSS in the SYN-ACK packet. Without this information, the data size in the SYN packet is limited to the default MSS of 536 bytes [RFC1122]. The client **SHOULD** update the cache MSS value whenever it discovers new MSS value, e.g., through path MTU discovery.

Caching RTT allows seeding a more accurate SYN timeout than the default value [RFC6298]. This lowers the performance penalty if the network or the server drops the SYN packets with data or the cookie options.

The cache replacement algorithm is not specified and is left to the implementations.

Note that before TFO sees wide deployment, clients **SHOULD** cache negative responses from servers in order to reduce the amount of futile TFO attempts. Since TFO is enabled on a per-service port basis but cookies are independent of service ports, clients' cache should include remote port numbers too.

4.2. Fast Open Protocol

One predominant requirement of TFO is to be fully compatible with existing TCP implementations, both on the client and the server sides.

The server keeps two variables per listening port:

FastOpenEnabled: default is off. It **MUST** be turned on explicitly by the application. When this flag is off, the server does not perform

any TFO related operations and MUST ignore all cookie options.

`PendingFastOpenRequests`: tracks number of TFO connections in SYN-RCVD state. If this variable goes over a preset system limit, the server SHOULD disable TFO for all new connection requests until `PendingFastOpenRequests` drops below the system limit. This variable is used for defending some vulnerabilities discussed in the "Security Considerations" section.

The server keeps a `FastOpened` flag per TCB to mark if a connection has successfully performed a TFO.

4.2.1. Fast Open Cookie Request

Any client attempting TFO MUST first request a cookie from the server with the following steps:

1. The client sends a SYN packet with a Fast Open Cookie Request option.
2. The server SHOULD respond with a SYN-ACK based on the procedures in the "Server Cookie Handling" section. This SYN-ACK SHOULD contain a Fast Open Cookie option if the server currently supports TFO for this listener port.
3. If the SYN-ACK contains a Fast Open Cookie option, the client replaces the cookie and other information as described in the "Client Cookie Handling" section. Otherwise, if the SYN-ACK is first seen, i.e., not a (spurious) retransmission, the client MAY remove the server information from the cookie cache. If the SYN-ACK is a spurious retransmission without valid Fast Open Cookie Option, the client does nothing to the cookie cache for the reasons below.

The network or servers may drop the SYN or SYN-ACK packets with the new cookie options, which will cause SYN or SYN-ACK timeouts. We RECOMMEND both the client and the server to retransmit SYN and SYN-ACK without the cookie options on timeouts. This ensures the connections of cookie requests will go through and lowers the latency penalty (of dropped SYN/SYN-ACK packets). The obvious downside for maximum compatibility is that any regular SYN drop will fail the cookie (although one can argue the delay in the data transmission till after 3WSH is justified if the SYN drop is due to network congestion). Next section describes a heuristic to detect such drops when the client receives the SYN-ACK.

We also RECOMMEND the client to record servers that failed to respond to cookie requests and only attempt another cookie request after

certain period. An alternate proposal is to request cookie in FIN instead since FIN-drop by incompatible middle-box does not affect latency. However such paths are likely to drop SYN packet with data later, and many applications close the connections with RST instead, so the actual benefit of this approach is not clear.

4.2.2. TCP Fast Open

Once the client obtains the cookie from the target server, it can perform subsequent TFO connections until the cookie is expired by the server.

Client: Sending SYN

To open a TFO connection, the client MUST have obtained a cookie from the server:

1. Send a SYN packet.
 - a. If the SYN packet does not have enough option space for the Fast Open Cookie option, abort TFO and fall back to regular 3WHS.
 - b. Otherwise, include the Fast Open Cookie option with the cookie of the server. Include any data up to the cached server MSS or default 536 bytes.
2. Advance to SYN-SENT state and update SND.NXT to include the data accordingly.
3. If RTT is available from the cache, seed SYN timer according to [RFC6298].

To deal with network or servers dropping SYN packets with payload or unknown options, when the SYN timer fires, the client SHOULD retransmit a SYN packet without data and Fast Open Cookie options.

Server: Receiving SYN and responding with SYN-ACK

Upon receiving the SYN packet with Fast Open Cookie option:

1. Initialize and reset a local FastOpened flag. If FastOpenEnabled is false, go to step 5.
2. If PendingFastOpenRequests is over the system limit, go to step 5.
3. If IsCookieValid() in section 4.1.2 returns false, go to step 5.
4. Buffer the data and notify the application. Set FastOpened flag

and increment PendingFastOpenRequests.

5. Send the SYN-ACK packet. The packet MAY include a Fast Open Option. If FastOpened flag is set, the packet acknowledges the SYN and data sequence. Otherwise it acknowledges only the SYN sequence. The server MAY include data in the SYN-ACK packet if the response data is readily available. Some application may favor delaying the SYN-ACK, allowing the application to process the request in order to produce a response, but this is left up to the implementation.
6. Advance to the SYN-RCVD state. If the FastOpened flag is set, the server MUST follow the congestion control [RFC5681], in particular the initial congestion window [RFC3390], to send more data packets.

Note that if SYN-ACK is lost, regular TCP reduces the initial congestion window before sending any data. In this case TFO is slightly more aggressive in the first data round trip even though it does not change the congestion control.

If the SYN-ACK timer fires, the server SHOULD retransmit a SYN-ACK segment with neither data nor Fast Open Cookie options for compatibility reasons.

A special case is simultaneous open where the SYN receiver is a client in SYN-SENT state. The protocol remains the same because [RFC793] already supports both data in SYN and simultaneous open. But the client's socket may have data available to read before it's connected. This document does not cover the corresponding API change.

Client: Receiving SYN-ACK

The client SHOULD perform the following steps upon receiving the SYN-ACK:

1. Update the cookie cache if the SYN-ACK has a Fast Open Cookie Option or MSS option or both.
2. Send an ACK packet. Set acknowledgment number to RCV.NXT and include the data after SND.UNA if data is available.
3. Advance to the ESTABLISHED state.

Note there is no latency penalty if the server does not acknowledge the data in the original SYN packet. The client SHOULD retransmit any unacknowledged data in the first ACK packet in step 2. The data exchange will start after the handshake like a regular TCP

connection.

If the client has timed out and retransmitted only regular SYN packets, it can heuristically detect paths that intentionally drop SYN with Fast Open option or data. If the SYN-ACK acknowledges only the initial sequence and does not carry a Fast Open cookie option, presumably it is triggered by a retransmitted (regular) SYN and the original SYN or the corresponding SYN-ACK was lost.

Server: Receiving ACK

Upon receiving an ACK acknowledging the SYN sequence, the server decrements PendingFastOpenRequests and advances to the ESTABLISHED state. No special handling is required further.

5. Security Considerations

The Fast Open cookie stops an attacker from trivially flooding spoofed SYN packets with data to burn server resources or to mount an amplified reflection attack on random hosts. The server can defend against spoofed SYN floods with invalid cookies using existing techniques [RFC4987]. We note that although generating bogus cookies is cost-free, the cost of validating the cookies, inherent to any authentication scheme, may not be substantial compared to processing a regular SYN packet.

5.1. Resource Exhaustion Attack by SYN Flood with Valid Cookies

However, the attacker may still obtain cookies from some compromised hosts, then flood spoofed SYN with data and "valid" cookies (from these hosts or other vantage points). With DHCP, it's possible to obtain cookies of past IP addresses without compromising any host. Below we identify new vulnerabilities of TFO and describe the countermeasures.

Like regular TCP handshakes, TFO is vulnerable to such an attack. But the potential damage can be much more severe. Besides causing temporary disruption to service ports under attack, it may exhaust server CPU and memory resources. Such an attack will show up on application server logs as a application level DoS from Bot-nets, triggering other defenses and alerts.

For this reason it is crucial for the TFO server to limit the maximum number of total pending TFO connection requests, i.e., PendingFastOpenRequests. When the limit is exceeded, the server temporarily disables TFO entirely as described in "Server Cookie Handling". Then subsequent TFO requests will be downgraded to regular connection requests, i.e., with the data dropped and only SYN

acknowledged. This allows regular SYN flood defense techniques [RFC4987] like SYN-cookies to kick in and prevent further service disruption.

The main impact of SYN floods against the standard TCP stack is not directly from the floods themselves costing TCP processing overhead or host memory, but rather from the spoofed SYN packets filling up the often small listener's queue.

On the other hand, TFO SYN floods can cause damage directly if admitted without limit into the stack. The RST packets from the spoofed host will fuel rather than defeat the SYN floods as compared to the non-TFO case, because the attacker can flood more SYNs with data to cost more data processing resources. For this reason, a TFO server needs to monitor the connections in SYN-RCVD being reset in addition to imposing a reasonable max queue length. Implementations may combine the two, e.g., by continuing to account for those connection requests that have just been reset against the listener's PendingFastOpenRequests until a timeout period has passed.

Limiting the maximum number of pending TFO connection requests does make it easy for an attacker to overflow the queue, causing TFO to be disabled. We argue that causing TFO to be disabled is unlikely to be of interest to attackers because the service will remain intact without TFO hence there is hardly any real damage.

5.1.1 Attacks from behind Sharing Public IPs (NATs)

An attacker behind NAT can easily obtain valid cookies to launch the above attack to hurt other clients that share the path. [BRISCOE12] suggested that the server can extend cookie generation to include the TCP timestamp---GetCookie(IP_Address, Timestamp)---and implement it by encrypting the concatenation of the two values to generate the cookie. The client stores both the cookie and its corresponding timestamp, and echoes both in the SYN. The server then implements IsCookieValid(IP_Address, Timestamp, Cookie) by encrypting the IP and timestamp data and comparing it with the cookie value.

This enables the server to issue different cookies to clients that share the same IP address, hence can selectively discard those misused cookies from the attacker. However the attacker can simply repeat the attack with new cookies. The server would eventually need to throttle all requests from the IP address just like the current approach. Moreover this approach requires modifying [RFC 1323] to send non-zero Timestamp Echo Reply in SYN, potentially cause firewall issues. Therefore we believe the benefit does not outweigh the drawbacks.

5.2. Amplified Reflection Attack to Random Host

Limiting PendingFastOpenRequests with a system limit can be done without Fast Open Cookies and would protect the server from resource exhaustion. It would also limit how much damage an attacker can cause through an amplified reflection attack from that server. However, it would still be vulnerable to an amplified reflection attack from a large number of servers. An attacker can easily cause damage by tricking many servers to respond with data packets at once to any spoofed victim IP address of choice.

With the use of Fast Open Cookies, the attacker would first have to steal a valid cookie from its target victim. This likely requires the attacker to compromise the victim host or network first.

The attacker here has little interest in mounting an attack on the victim host that has already been compromised. But it may be motivated to disrupt the victim's network. Since a stolen cookie is only valid for a single server, it has to steal valid cookies from a large number of servers and use them before they expire to cause sufficient damage without triggering the defense.

One can argue that if the attacker has compromised the target network or hosts, it could perform a similar but simpler attack by injecting bits directly. The degree of damage will be identical, but TFO-specific attack allows the attacker to remain anonymous and disguises the attack as from other servers.

The best defense is for the server not to respond with data until handshake finishes. In this case the risk of amplification reflection attack is completely eliminated. But the potential latency saving from TFO may diminish if the server application produces responses earlier before the handshake completes.

6. TFO's Applicability

This section is to help applications considering TFO to evaluate TFO's benefits and drawbacks using the Web client and server applications as an example throughout. A proposed socket API change is in the Appendix.

6.1 Duplicate Data in SYNs

It is possible, though uncommon, that using TFO results in the first data written to a socket to be delivered more than once to the application on the remote host (Section 2.1). This replay potential only applies to data in the SYN but not subsequent data exchanges. The client **MUST NOT** use TFO to send data in the SYN, and the server **MUST NOT** accept data in the SYN if it cannot handle receiving the same SYN data more than once, due to reasons described before.

6.2 Potential Performance Improvement

TFO is designed for latency-conscious applications that are sensitive to TCP's initial connection setup delay. To benefit from TFO, the first application data unit (e.g., an HTTP request) needs to be no more than TCP's maximum segment size (minus options used in SYN). Otherwise the remote server can only process the client's application data unit once the rest of it is delivered after the initial handshake, diminishing TFO's benefit.

To the extent possible, applications **SHOULD** reuse the connection to take advantage of TCP's built-in congestion control and reduce connection setup overhead. An application employs too many short-lived connections will negatively impact network stability, as these connections often exit before TCP's congestion control algorithm takes effect.

6.3. Example: Web Clients and Servers

6.3.1. HTTP Request Replay

While TFO is motivated by Web applications, the browser should not use TFO to send requests in SYNs if those requests cannot tolerate replays. One example is POST requests without application-layer transaction protection (e.g., a unique identifier in the request header).

TFO is particularly useful for GET requests. Even though not all GET requests are idempotent, GETs are frequently replayed today across striped TCP connections. After a server receives an HTTP request but before the ACKs of the requests reach the browser, the browser may

timeout and retry the same request on another (possibly new) TCP connection. This differs from a TFO replay only in that the replay is initiated by the browser, not by the TCP stack.

Finally, TFO is safe and useful for HTTPS requests because it saves the first SSL handshake RTT and the HTTP request is sent after the connection establishes.

6.3.2. Comparison with HTTP Persistent Connections

Is TFO useful given the wide deployment of HTTP persistent connections? The short answer is yes. Studies [RCCJR11][AERG11] show that the average number of transactions per connection is between 2 and 4, based on large-scale measurements from both servers and clients. In these studies, the servers and clients both kept idle connections up to several minutes, well into "human think" time.

Keeping connections open and idle even longer risks a greater performance penalty. [HNESSK10][MQXMZ11] show that the majority of home routers and ISPs fail to meet the the 124-minute idle timeout mandated in [RFC5382]. In [MQXMZ11], 35% of mobile ISPs silently timeout idle connections within 30 minutes. End hosts, unaware of silent middle-box timeouts, suffer multi-minute TCP timeouts upon using those long-idle connections.

To circumvent this problem, some applications send frequent TCP keep-alive probes. However, this technique drains power on mobile devices [MQXMZ11]. In fact, power has become such a prominent issue in modern LTE devices that mobile browsers close HTTP connections within seconds or even immediately [SOUDERS11].

[RCCJR11] studied Chrome browser performance based on 28 days of global statistics. The Chrome browser keeps idle HTTP persistent connections for 5 to 10 minutes. However the average number of the transactions per connection is only 3.3 and TCP 3WHS accounts for up to 25% of the HTTP transaction network latency. The authors tested a Linux TFO implementation with TFO enabled Chrome browser on popular web sites in emulated environments such as residential broadband and mobile networks. They showed that TFO improves page load time by 10% to 40%.

7. Open Areas for Experimentation

We now outline some areas that need experimentation in the Internet and under different network scenarios. These experiments should help the community evaluate Fast Open benefits and risks towards further standardization and implementation of Fast Open and its related protocols.

7.1. Performance impact due to middle-boxes and NAT

[MAF04] found that some middle-boxes and end-hosts may drop packets with unknown TCP options. Studies [LANGLEY06, HNRGHT11] both found that 6% of the probed paths on the Internet drop SYN packets with data or with unknown TCP options. The TFO protocol deals with this problem by falling back to regular TCP handshake and re-transmitting SYN without data or cookie options after the initial SYN timeout. Moreover the implementation is recommended to negatively cache such incidents to avoid recurring timeouts. Further study is required to evaluate the performance impact of these malicious drop behaviors.

Another interesting study is the (loss of) TFO performance benefit behind certain carrier-grade NAT. Typically hosts behind a NAT sharing the same IP address will get the same cookie for the same server. This will not prevent TFO from working. But on some carrier-grade NAT configurations where every new TCP connection from the same physical host uses a different public IP address, TFO does not provide latency benefits. However, there is no performance penalty either, as described in Section "Client: Receiving SYN-ACK".

7.2. Cookie-less Fast Open

The cookie mechanism mitigates resource exhaustion and amplification attacks. However cookies are not necessary if the server has application-level protection or is immune to these attacks. For example a Web server that only replies with a simple HTTP redirect response that fits in the SYN-ACK packet may not care about resource exhaustion. Such an application can safely disable TFO cookie checks.

Disabling cookies simplifies both the client and the server, as the client no longer needs to request a cookie and the server no longer needs to check or generate cookies. Disabling cookies also potentially simplifies configuration, as the server no longer needs a key. It may be preferable to enable SYN cookies and disable TFO [RFC4987] when a server is overloaded by a large-scale Bot-net attack.

Careful experimentation is necessary to evaluate if cookie-less TFO is practical. The implementation can provide an experimental feature to allow zero length, or null, cookies as opposed to the minimum 4 bytes cookies. Thus the server may return a null cookie and the client will send data in SYN with it subsequently. If the server believes it's under a DoS attack through other defense mechanisms, it can switch to regular Fast Open for listener sockets.

8. Related Work

8.1. T/TCP

TCP Extensions for Transactions [RFC1644] attempted to bypass the three-way handshake, among other things, hence shared the same goal but also the same set of issues as TFO. It focused most of its effort battling old or duplicate SYNs, but paid no attention to security vulnerabilities it introduced when bypassing 3WS [PHRACK98].

As stated earlier, we take a practical approach to focus TFO on the security aspect, while allowing old, duplicate SYN packets with data after recognizing that 100% TCP semantics is likely infeasible. We believe this approach strikes the right tradeoff, and makes TFO much simpler and more appealing to TCP implementers and users.

8.2. Common Defenses Against SYN Flood Attacks

[RFC4987] studies on mitigating attacks from regular SYN flood, i.e., SYN without data. But from the stateless SYN-cookies to the stateful SYN Cache, none can preserve data sent with SYN safely while still providing an effective defense.

The best defense may be to simply disable TFO when a host is suspected to be under a SYN flood attack, e.g., the SYN backlog is filled. Once TFO is disabled, normal SYN flood defenses can be applied. The "Security Consideration" section contains a thorough discussion on this topic.

8.3. TCP Cookie Transaction (TCPCT)

TCPCT [RFC6013] eliminates server state during initial handshake and defends spoofing DoS attacks. Like TFO, TCPCT allows SYN and SYN-ACK packets to carry data. But the server can only send up to MSS bytes of data during the handshake instead of the initial congestion window unlike TFO. Therefore applications like Web may not receive the latency benefit as TFO.

8.4. Speculative Connections by the Applications

Some Web browsers maintain a history of the domains for frequently visited web pages. The browsers then speculatively pre-open TCP connections to these domains before the user initiates any requests for them [BELSHELL]. The downside of this approach is that it wastes server and network resources by initiating and maintaining idle connections; It is also subject to the NAT timeout issues described in Section 6.3.2. TFO offers similar performance improvement without the added overhead.

9. IANA Considerations

The Fast Open Cookie Option and Fast Open Cookie Request Option define no new namespace. The options require IANA to allocate one value from the TCP option Kind namespace. Early implementation before the IANA allocation SHOULD follow [EXPOPT] and use experimental option 254 and magic number 0xF989 (16 bits), then migrate to the new option after the allocation accordingly.

10. Acknowledgement

We thank Rick Jones, Bob Briscoe, Adam Langley, Matt Mathis, Neal Cardwell, Roberto Peon, William Chan, Eric Dumazet, and Tom Herbert for their feedbacks. We especially thank Barath Raghavan for his contribution on the security design of Fast Open and proofreading this draft numerous times.

11. References

11.1. Normative References

- [RFC793] Postel, J. "Transmission Control Protocol", RFC 793, September 1981.
- [RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.
- [RFC5382] S. Guha, Ed., Biswas, K., Ford B., Sivakumar S., Srisuresh, P., "NAT Behavioral Requirements for TCP", RFC 5382
- [RFC5681] Allman, M., Paxson, V. and E. Blanton, "TCP Congestion Control", RFC 5681, September 2009.
- [RFC6298] Paxson, V., Allman, M., Chu, J. and M. Sargent, "Computing TCP's Retransmission Timer", RFC 6298, June 2011.
- [RFC6928] Chu, J., Dukkkipati, N., Cheng, Y. and M. Mathis, "Increasing TCP's Initial Window", RFC 6928, April 2013.

11.2. Informative References

- [AERG11] M. Al-Fares, K. Elmeleegy, B. Reed, and I. Gashinsky, "Overclocking the Yahoo! CDN for Faster Web Page Loads". In Proceedings of Internet Measurement Conference, November 2011.
- [EXPOPT] Touch, Joe, "Shared Use of Experimental TCP Options", Internet-Draft draft-ietf-tcpm-experimental-options (work in progress), June 2013.
- [HNESSK10] S. Haetoenen, A. Nyrhinen, L. Eggert, S. Strowes, P.

Sarolahti, M. Kojo., "An Experimental Study of Home Gateway Characteristics". In Proceedings of Internet Measurement Conference. October 2010

- [HNRGHT11] M. Honda, Y. Nishida, C. Raiciu, A. Greenhalgh, M. Handley, H. Tokuda, "Is it Still Possible to Extend TCP?". In Proceedings of Internet Measurement Conference. November 2011.
- [LANGLEY06] Langley, A, "Probing the viability of TCP extensions", URL <http://www.imperialviolet.org/binary/ecntest.pdf>
- [MAF04] Medina, A., Allman, M., and S. Floyd, "Measuring Interactions Between Transport Protocols and Middleboxes", In Proceedings of Internet Measurement Conference, October 2004.
- [MQXMZ11] Z. Mao, Z. Qian, Q. Xu, Z. Mao, M. Zhang. "An Untold Story of Middleboxes in Cellular Networks", In Proceedings of SIGCOMM. August 2011.
- [PHRACK98] "T/TCP vulnerabilities", Phrack Magazine, Volume 8, Issue 53 artical 6. July 8, 1998. URL <http://www.phrack.com/issues.html?issue=53&id=6>
- [QWGMSS11] F. Qian, Z. Wang, A. Gerber, Z. Mao, S. Sen, O. Spatscheck. "Profiling Resource Usage for Mobile Applications: A Cross-layer Approach", In Proceedings of International Conference on Mobile Systems. April 2011.
- [RCCJR11] Radhakrishnan, S., Cheng, Y., Chu, J., Jain, A. and Raghavan, B., "TCP Fast Open". In Proceedings of 7th ACM CoNEXT Conference, December 2011.
- [RFC1644] Braden, R., "T/TCP -- TCP Extensions for Transactions Functional Specification", RFC 1644, July 1994.
- [RFC4987] Eddy, W., "TCP SYN Flooding Attacks and Common Mitigations", RFC 4987, August 2007.
- [RFC6013] Simpson, W., "TCP Cookie Transactions (TCPCT)", RFC6013, January 2011.
- [SOUDERS11] S. Souders. "Making A Mobile Connection". <http://www.stevesouders.com/blog/2011/09/21/making-a-mobile-connection/>
- [THK98] Touch, J., Heidemann, J., Obraczka, K., "Analysis of HTTP

Performance", USC/ISI Research Report 98-463. December 1998.

[BRISCOE12] Briscoe, B., "Some ideas building on draft-ietf-tcpm-fastopen-01", tcpm list, [http://www.ietf.org/mail-archive/web/tcpm/current/January 16, 2014msg07192.html](http://www.ietf.org/mail-archive/web/tcpm/current/January%2016,%202014msg07192.html)

[BELSHE12] Belshe, M., "The era of browser preconnect.", <http://www.belshe.com/2011/02/10/the-era-of-browser-preconnect/>

Appendix A. Example Socket API Changes to support TFO

The design rationale is to minimize changes to the socket API and hence applications, in order to reduce the deployment hurdle. The following changes have been implemented in Linux 3.7 or later kernels.

A.1 MSG_FASTOPEN flag for sendto() or sendmsg()

MSG_FASTOPEN marks the attempt to send data in SYN like a combination of connect() and sendto(), by performing an implicit connect() operation. It blocks until the handshake has completed and the data is buffered.

For non-blocking socket it returns the number of bytes buffered and sent in the SYN packet. If the cookie is not available locally, it returns -1 with errno EINPROGRESS, and sends a SYN with TFO cookie request automatically. The caller needs to write the data again when the socket is connected.

It returns the same errno as connect() if the handshake fails.

A.2 TCP_FASTOPEN setsockopt() Socket Option

The option enables Fast Open on the listener socket. The option value specifies the PendingFastOpenRequests threshold, i.e., the maximum length of pending SYNs with data payload. Once enabled, the TCP implementation will respond with TFO cookies per request.

Previously accept() returns only after a socket is connected. But for a Fast Open connection, accept() returns upon receiving a SYN with a valid Fast Open cookie and data, and the data is available to be read through, e.g., recvmsg(), read().

Authors' Addresses

Yuchung Cheng Google, Inc. 1600 Amphitheatre Parkway Mountain View,
CA 94043, USA EMail: ycheng@google.com

Jerry Chu Google, Inc. 1600 Amphitheatre Parkway Mountain View, CA
94043, USA EMail: hkchu@google.com

Sivasankar Radhakrishnan Department of Computer Science and
Engineering University of California, San Diego 9500 Gilman Dr La
Jolla, CA 92093-0404 EMail: sivasankar@cs.ucsd.edu

Arvind Jain Google, Inc. 1600 Amphitheatre Parkway Mountain View, CA
94043, USA EMail: arvind@google.com