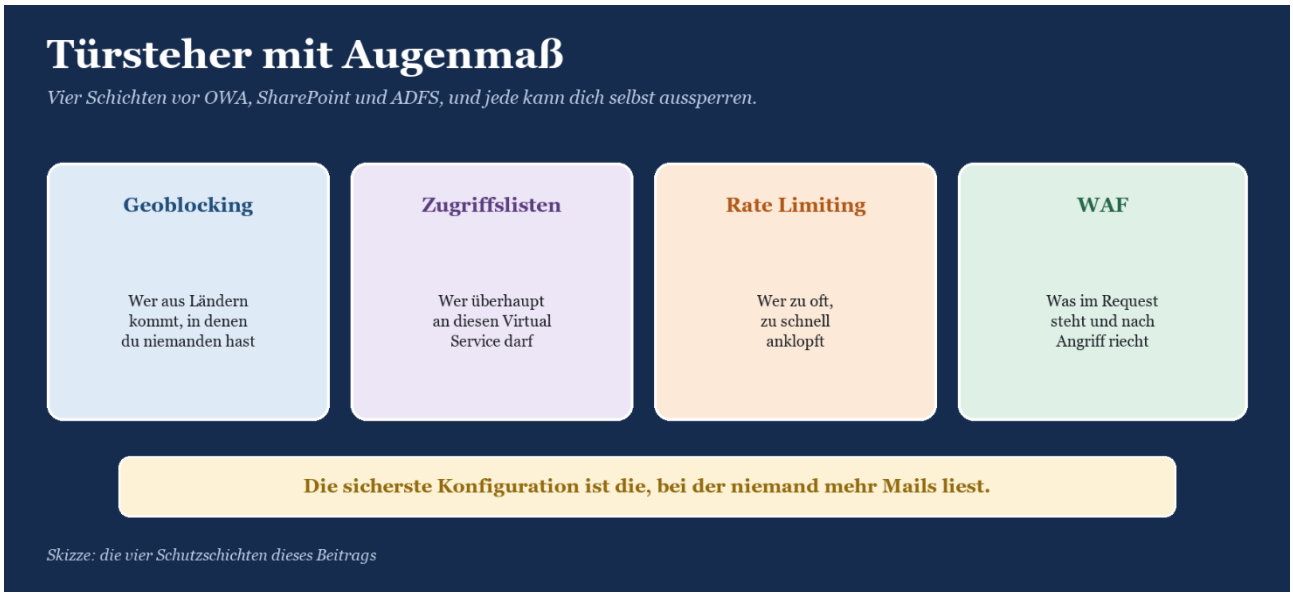
Web Application Firewall und Geoblocking am LoadMaster: Schutz ohne Fehlalarm-Hölle



WISSEN

Alles zum Kemp LoadMaster mit Microsoft-Brille: Exchange, SharePoint, ADFS, RDS, Eigenentwicklungen und Kerberos an einem Ort.

› **Kemp LoadMaster**

BERATUNG

Kemp WAF, Geoblocking und Rate Limiting so einführen, dass Angreifer draußen bleiben und deine Anwenderinnen und Anwender nicht gleich mit?

› **Beratung zum Kemp LoadMaster**

SCHULUNG

Regel-IDs lesen, Ausnahmen bauen, Schwellen drehen: im Labor üben, bevor das produktive OWA zum Versuchskaninchen wird.

› **Kemp-LoadMaster-Schulung**

Es gibt zwei Sorten von Administratoren, die eine Web Application Firewall betreiben. Die einen haben sie nie eingeschaltet, weil beim ersten Versuch Outlook im Web zusammengebrochen ist. Die anderen haben sie eingeschaltet, im Protokollmodus vergessen und zeigen seit drei Jahren in jedem Audit stolz auf das grüne Häkchen. Beide Gruppen schlafen gut. Beide Gruppen sind nicht geschützt. Dazwischen liegt ein schmaler Grat, auf dem die WAF tatsächlich Angriffe blockiert und das Team trotzdem noch E-Mails mit Anhängen verschicken kann. Genau um diesen Grat geht es hier.

Die Kemp WAF am LoadMaster ist dabei kein exotisches Spezialwerkzeug, sondern ModSecurity mit dem OWASP Core Rule Set unter einer Bedienoberfläche, die man auch nachts um drei noch versteht. Das Problem ist selten das Werkzeug. Das Problem ist, dass Exchange, SharePoint und ADFS Dinge tun, die für ein generisches Regelwerk verdächtig aussehen: riesige POST-Bodies, Base64-kodierte Anhänge, JSON in HTTP-Headern, SOAP-Umschläge und Binärprotokolle. Das Core Rule Set kennt WordPress, Drupal und Nextcloud mit eigenen Ausnahmepaketen. Für Exchange und SharePoint gibt es so etwas nicht. Die Ausnahmen baust du selbst, und wer das ohne Plan tut, landet in der Fehlalarm-Hölle.

Dazu kommen Geoblocking, Zugriffslisten je Virtual Service und Rate Limiting. Alle drei klingen nach dem perfekten Rezept gegen Passwort-Spraying auf OWA und ADFS, und alle drei haben Grenzen, über die in Herstellerunterlagen eher leise gesprochen wird: VPN-Ausgänge, Mobilfunk-NAT, Microsoft-365-Hybridverkehr und die Vertriebskollegin, die aus dem Hotel in Singapur ihre

Mails lesen will. Dieser Beitrag gehört zur Serie rund um den Kemp LoadMaster und schließt die Lücke zwischen Progress-Dokumentation und Microsoft Learn: warum die Dinge so sind, wie sie sind, und wo es in der Praxis knallt.

So tickt die Kemp WAF: Regeln, Punkte und Paranoia

Die WAF des LoadMasters arbeitet regelbasiert. Sie vergleicht jeden Request, auf Wunsch auch jede Antwort, mit einem Regelwerk und entscheidet anhand der Treffer, ob der Verkehr durchgeht. Seit LoadMaster-Firmware 7.2.54 mit dem WAF-Stand 1.5 ist das OWASP Core Rule Set, kurz CRS, nativ eingebunden, inklusive Anomaly Scoring und Paranoia Levels. Ältere Konfigurationen mit den früheren kommerziellen Regelpaketen tauchen in der Oberfläche als Legacy-WAF auf. Wer noch dort steht, sollte die Migration auf das CRS einplanen, denn die Optionen und Begriffe unterscheiden sich, und Anleitungen aus dem Netz passen sonst nicht zu dem, was du siehst.

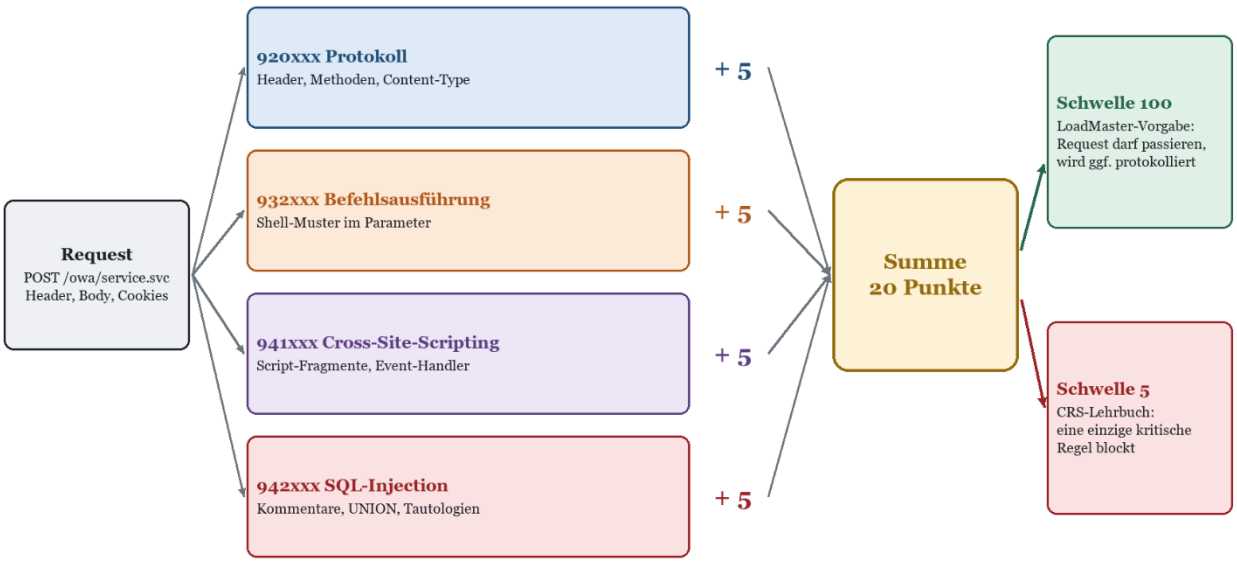
Anomaly Scoring statt Sofortschuss

Der wichtigste Mechanismus, den du verstehen musst, ist das Anomaly Scoring. Eine einzelne Regel blockiert im CRS normalerweise nicht selbst. Sie legt Punkte auf einen Stapel: kritische Treffer bringen fünf Punkte, Fehler vier, Warnungen drei, Hinweise zwei. Am Ende der Request-Phase prüft eine Sammelregel, ob die Summe die eingestellte Schwelle erreicht. Erst dann wird geblockt. Im Audit-Log erkennst du das an der Regel 949110 mit der Meldung „Inbound Anomaly Score Exceeded“ und der Gesamtpunktzahl.

Und jetzt der Punkt, an dem viele stolpern: Der LoadMaster setzt den Anomaly Scoring Threshold standardmäßig auf 100. Das CRS-Lehrbuch empfiehlt für den scharfen Betrieb 5, also einen einzigen kritischen Treffer. Bei 100 müssten rund zwanzig kritische Regeln gleichzeitig anschlagen. Das passiert bei plumpen Scannern, aber kaum bei einem gezielten Angriff. Die Vorgabe ist also bewusst nachsichtig, damit nach dem Einschalten nicht sofort alles brennt. Wer sie nie anfasst, hat eine WAF, die vor allem Logdateien produziert.

Anomaly Scoring: die WAF zählt mit, statt sofort zu schießen

Jede Regel, die anschlägt, legt Punkte auf den Stapel; geblockt wird erst an der Schwelle



Merke: Die Schwelle ist dein Lautstärkeregler. Von oben nach unten drehen, nie umgekehrt.

Skizze: vereinfacht; Punktwerte gelten für Regeln mit Schweregrad kritisch

Paranoia Level: wie misstrauisch darf es sein?

Das CRS teilt seine Regeln in vier Paranoia Levels ein. Stufe 1 enthält Regeln mit wenig Fehlalarmen, Stufe 4 ist für Umgebungen gedacht, in denen jemand dafür bezahlt wird, den ganzen Tag Ausnahmen zu pflegen. Der LoadMaster trennt dabei zwei Werte: Der Blocking Paranoia Level bestimmt, welche Regeln zur blockierenden Punktzahl beitragen, der Executing Paranoia Level, welche Regeln zusätzlich ausgeführt und nur protokolliert werden. Der zweite Wert darf nicht unter dem ersten liegen. Das ist ein geniales Werkzeug für die Lernphase: Du blockst mit Stufe 1 und schaust dir mit Stufe 2 an, was passieren würde, wenn du strenger wärst.

Einstellung je Virtual Service	Vorgabe	Was sie bewirkt	Microsoft-Brille
Anomaly Scoring Threshold	100	Punktsumme, ab der geblockt wird (1 bis 10000)	Schrittweise senken, für OWA landen viele zwischen 5 und 20
Blocking Paranoia Level	1	Strenge der blockierenden Regeln	Für Exchange und SharePoint bei 1 bleiben, alles darüber ist Hobby
Executing Paranoia Level	nicht kleiner als Blocking	Zusätzliche Regeln nur protokollieren	In der Lernphase eine Stufe höher als Blocking
Audit mode	je nach Stand	No Audit, Audit Relevant oder Audit All	Audit Relevant im Betrieb, Audit All nur kurz zur Fehlersuche
Inspect HTTP POST Request Bodies	aus	Bodies inklusive JSON und XML prüfen	Ohne das sieht die WAF bei OWA und SharePoint die Hälfte nicht
Request Body Size Limit	1 MB	Maximale Body-Größe, höchstens 50 MB	Kollidiert direkt mit Exchange-Anhängen und SharePoint-Uploads
Process HTTP Responses	je nach Stand	Antworten prüfen, kostet CPU und Speicher	Für OWA selten lohnend, für Eigenentwicklungen interessant
Countries to block	keine	Länderbasierte Sperre direkt in der WAF	Siehe Abschnitt Geoblocking, Hybridverkehr beachten

FAKTEN · Kemp WAF in Zahlen

Regelwerk: OWASP Core Rule Set auf Basis von ModSecurity, nativ seit Firmware 7.2.54 (WAF 1.5).

Anomaly Scoring Threshold: Vorgabe 100, zulässig 1 bis 10000; die meisten kritischen Regeln zählen 5 Punkte.

Request Body Size Limit: Vorgabe 1.048.576 Byte, höchstens 52.428.800 Byte, also 50 MB.

Konfiguration: je Virtual Service oder SubVS, nicht gleichzeitig auf beiden Ebenen. Die WAF braucht entschlüsselten Verkehr, also TLS-Offload oder Re-Encryption am LoadMaster.

Lizenz: WAF und Geo-Funktionen sind abonnementpflichtig; graue Optionen in der Oberfläche bedeuten fast immer eine Lizenzfrage, keinen Bug.

Der letzte Punkt im Kasten ist wichtiger, als er klingt: Eine WAF, die nur verschlüsselte Bytes sieht, prüft nichts. Sie muss den Verkehr also am LoadMaster entschlüsseln, und damit hängen Zertifikate, Cipher-Auswahl und die Frage nach Re-Encryption zum Exchange direkt an diesem Thema. Wie du das sauber aufsetzt, beschreibt der Beitrag TLS-Offload, Re-Encryption und Zertifikate: PKI trifft LoadMaster. Welche Edition die WAF überhaupt mitbringt, klärt Editionen, Lizenzierung und Sizing: welcher LoadMaster für welchen Laden.

Lernphase, Blockieren und das Audit-Log lesen

Eine WAF wird nicht eingeschaltet, sie wird eingeführt. Das klingt nach Beraterlyrik, ist aber der Unterschied zwischen einem ruhigen Montag und einem Montag, an dem die Geschäftsführung persönlich im Serverraum steht. Der Ablauf ist immer derselbe: beobachten, Ausnahmen bauen, grob scharf schalten, fein nachziehen.

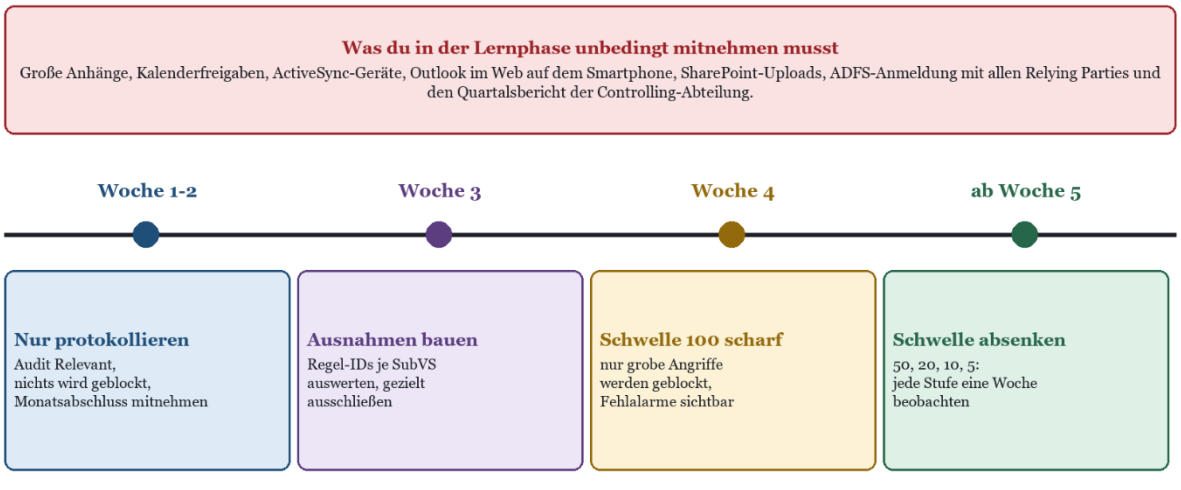
Die Lernphase richtig aufsetzen

In der Lernphase soll die WAF alles sehen und nichts blocken. In der Legacy-WAF gab es dafür einen eigenen Schalter „Audit Only“. Beim CRS-basierten WAF erreichst du denselben Effekt, indem du Audit Relevant aktivierst und die Schwelle so hoch setzt, dass praktisch nichts blockiert, gegebenenfalls bis zum Maximum. Dazu den Executing Paranoia Level eine Stufe über dem Blocking Paranoia Level und die Body-Inspektion an. Dann wartest du. Nicht einen Nachmittag, sondern so lange, bis alle typischen Vorgänge mindestens einmal vorbeigekommen sind.

- **Zeitraum:** zwei Wochen sind ein gutes Minimum, ein Monatsabschluss gehört idealerweise hinein, weil dann Leute Dinge tun, die sie sonst nie tun.
- **Clients:** Outlook im Web am Desktop und am Smartphone, Outlook klassisch und neu, ActiveSync-Geräte, Teams-Kalenderintegration, Scanner, die per EWS Mails verschicken.
- **Funktionen:** große Anhänge, Kalenderfreigaben, Abwesenheitsnotizen mit Sonderzeichen, Signaturen mit eingebetteten Bildern, SharePoint-Uploads über Browser und Explorer-Ansicht.
- **Identität:** ADFS-Anmeldungen für alle Relying Parties, auch die, an die sich niemand erinnert, bis die Reisekostenabrechnung nicht mehr geht.

Von der Lernphase zum Blockieren: ein realistischer Fahrplan

Wochen statt Minuten, und jede Stufe endet mit einem Blick ins Audit-Log



Rückweg einplanen: Jede Stufe muss in einer Minute zurückgedreht werden können.

Skizze: Erfahrungswerte aus Projekten, keine Herstellervorgabe

Scharf schalten in Stufen

Nach der Lernphase schaltest du mit der Vorgabe 100 scharf. Damit blockst du Scanner und plumpe Angriffe, aber kaum Fehlalarme. Danach senkst du die Schwelle in Stufen, etwa auf 50, 20, 10 und schließlich 5, und gibst jeder Stufe mindestens eine Woche. Sinkt die Schwelle, tauchen neue Fehlalarme auf, die du gezielt ausschließt, bevor du weiter drehst. Der Weg führt immer von oben nach unten. Wer mit 5 startet und dann hochdreht, weil alles kaputt ist, hat keine Lernphase, sondern einen Notfall.

WARNUNG · Nicht alle SubVS gleichzeitig scharf schalten

Wenn du die Schwelle für OWA, EWS, ECP und SharePoint am selben Morgen senkst, weißt du bei der ersten Störung nicht, wo du suchen sollst. Ein SubVS pro Woche, beginnend mit dem, das am wenigsten Nutzer hat. ECP ist dafür ideal: Wenn es dort klemmt, merken es ein paar Admins, nicht die ganze Belegschaft.

Audit-Logs lesen, ohne zu verzweifeln

Die WAF protokolliert im Format von ModSecurity. Für jeden auffälligen Request gibt es einen Eintrag mit eindeutiger ID, Zeitstempel, Client-Adresse, Pfad und einer Liste der Regeln, die angeschlagen haben. Welche Abschnitte im Log landen, steuerst du über die Audit Parts; B steht für die Request-Header, E für den Antwort-Body, F für die Antwort-Header und H für den Abschlussblock mit den Regelmeldungen. Der Abschlussblock ist der, den du wirklich liest.

Jede Regelmeldung enthält die Regel-ID, eine Klartextbeschreibung, den auslösenden Teil des Requests, etwa ARGS:body oder REQUEST_HEADERS:Cookie, und die betroffene URI. In der Logansicht des LoadMasters kannst du dir die Regel anzeigen lassen und sie auch direkt deaktivieren. Das ist bequem und gefährlich zugleich, denn ein Klick dort schaltet die Regel für den ganzen Virtual Service ab, nicht nur für den einen Pfad, der Ärger macht.

Regel-ID-Bereich	Thema	Typischer Fehlalarm bei Microsoft-Diensten
200002, 200003	Body nicht lesbar, Multipart fehlerhaft	Binärprotokolle, ActiveSync, abgebrochene Uploads
920xxx	Protokoll-Einhaltung	Ungewöhnliche Content-Types, lange Header, numerischer Host bei Health Checks
921xxx	Protokoll-Angriffe	Mehrteilige Requests von Office-Clients
930xxx, 931xxx	Datei- und Pfadangriffe	Pfadangaben in SharePoint-Parametern
932xxx	Befehlsausführung	Freitext in Mails und Kalendereinträgen, Skriptbeispiele
941xxx	Cross-Site-Scripting	HTML-Mails, Signaturen, Rich-Text-Felder
942xxx	SQL-Injection	Kommentarzeichen, Anführungszeichen, Wörter wie select in Freitext
949110	Sammelregel eingehend	Kein Fehlalarm, sondern das Ergebnis: Summe erreicht
959100	Sammelregel ausgehend	Nur mit Antwortprüfung relevant

TIPP · So arbeitest du ein Log systematisch ab

Exportiere die Einträge eines Tages und gruppiere nach Regel-ID und Pfad. Fünf Regeln erzeugen typischerweise den Großteil der Treffer. Für jede beantwortest du drei Fragen: Ist es legitimer Verkehr? Betrifft es nur einen Pfad oder einen Parameter? Gibt es eine engere Ausnahme als das Abschalten der ganzen Regel? Erst wenn die dritte Antwort Nein lautet, deaktivierst du eine Regel komplett.

Schick die WAF-Ereignisse zusätzlich per Syslog an dein zentrales Logging, damit du Blockaden mit Helpdesk-Tickets und Anmeldeereignissen korrelieren kannst; wie das geht, steht in Monitoring, Syslog und SNMP: den LoadMaster ins Monitoring holen.

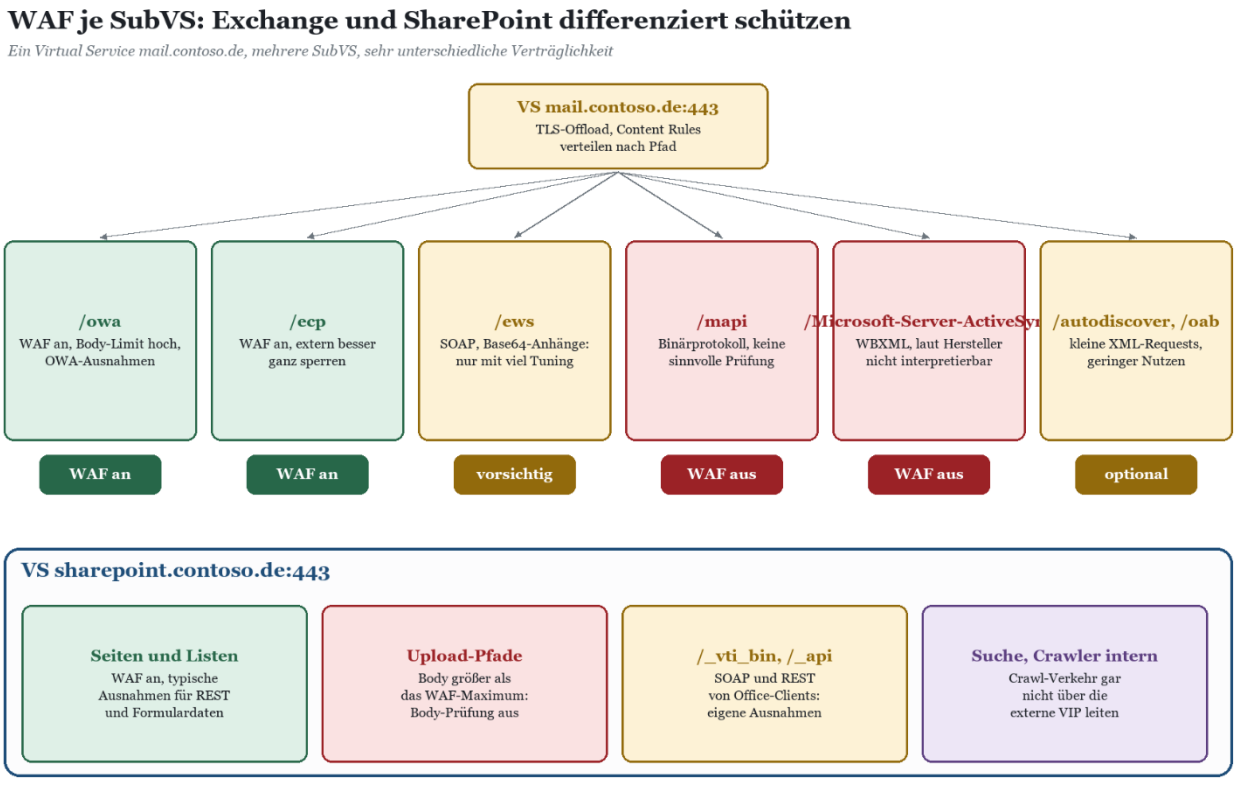
Ausnahmen für Exchange und SharePoint: große Uploads, Base64, Web-Services

Jetzt wird es konkret. Exchange und SharePoint sind keine Webseiten, sie sind Anwendungsplattformen, die zufällig HTTP sprechen. Ein generisches Regelwerk sieht dort

Muster, die in einem Webshop tatsächlich verdächtig wären: Base64-Blöcke in Megabyte-Größe, XML-Umschläge mit verschachtelten Elementen, Freitext mit Anführungszeichen und spitzen Klammern. Die Kunst besteht darin, Ausnahmen so eng zu schneiden, dass der Schutz für den Rest erhalten bleibt.

Erste Regel: WAF auf SubVS-Ebene, nicht pauschal

Progress empfiehlt selbst, die WAF bei Exchange auf SubVS-Ebene zu aktivieren, weil sie die Protokolle von ActiveSync nicht interpretieren kann. Das ist die halbe Miete. Die Exchange-Templates des LoadMasters bauen ohnehin einen Virtual Service mit SubVS je Dienst, die per Content Rule nach Pfad verteilt werden. Genau diese Struktur nutzt du: WAF an, wo Browser sprechen, WAF aus, wo Binärprotokolle laufen, und dazwischen sorgfältig. Wie die SubVS-Struktur aufgebaut ist, erklären Virtual Service, Real Server, SubVS: das Grundvokabular des LoadMasters und Exchange Server hinter dem LoadMaster: Templates, Namespaces, Health Checks.



Skizze: Ausgangspunkt für eigene Tests, Pfade vereinfacht, Namen fiktiv

Exchange-Dienst	Inhalt	WAF-Empfehlung	Ausgleich, wenn WAF aus
Outlook im Web (/owa)	HTML, JSON-Aktionen, Anhänge als Upload	An, mit Body-Prüfung und angehobenem Body-Limit	nicht nötig
Exchange Admin Center (/ecp)	Verwaltungsoberfläche	An, extern am besten ganz sperren	Zugriffsliste nur interne Netze
Exchange Web Services (/ews)	SOAP-XML, Anhänge als Base64	Vorsichtig, nur mit Ausnahmen für Body-Parameter	Rate Limiting, MFA davor wo möglich
MAPI over HTTP (/mapi)	Binäre Remote-Operationen	Aus	Rate Limiting, Zugriffsliste
ActiveSync	WBXML, binär	Aus, laut Hersteller nicht interpretierbar	Rate Limiting, Geräteverwaltung in

			Exchange
Autodiscover, OAB	Kleine XML-Requests, Downloads	Optional, geringer Nutzen	nicht nötig

Große Uploads: das 1-MB-Problem

Das Request Body Size Limit steht ab Werk auf 1 MB. Exchange erlaubt in der Standardkonfiguration Nachrichten von rund 25 MB in der Transportschicht und für Outlook im Web noch etwas mehr, weil Anhänge beim Hochladen Base64-kodiert werden und dabei etwa ein Drittel an Größe zulegen. Ein Anhang mit 1 MB auf der Platte ist im Request also schon deutlich größer als 1 MB. Mit Vorgabe-Limit scheitert damit jeder zweite Anhang, und die Fehlermeldung in Outlook im Web verrät natürlich nicht, dass eine WAF beteiligt war.

Für das OWA-SubVS hebst du das Limit also an, und zwar passend zu dem, was Exchange selbst zulässt, plus Kodierungsaufschlag. SharePoint ist härter: Dort sind Uploads von mehreren hundert Megabyte normal, und das WAF-Maximum liegt bei 50 MB. Für Upload-Pfade bleibt dann nur, die Body-Prüfung für genau diese Pfade auszusetzen oder sie in ein eigenes SubVS ohne Body-Inspektion zu legen. Das ist kein Schönheitsfehler, sondern eine bewusste Abwägung: Ein Dateiupload ist kein Formular, und die interessanten Angriffe gegen SharePoint zielen ohnehin auf andere Endpunkte. Was SharePoint hinter dem LoadMaster sonst noch braucht, steht in SharePoint Server Subscription Edition hinter dem LoadMaster; die Plattform selbst behandelt die Seite zu SharePoint.

WICHTIG · Limits müssen auf allen Ebenen zusammenpassen

Exchange-Transportlimits, die Limits in der web.config von OWA und EWS, das Request Body Size Limit der WAF und gegebenenfalls Grenzen am vorgelagerten Firewall-System müssen zueinander passen. Ist die WAF der kleinste Wert, blockt sie. Ist sie der größte, ist das in Ordnung, denn dann entscheidet Exchange mit einer verständlichen Fehlermeldung.

Base64, JSON in Headern und Freitext

Base64-Blöcke sind für Regeln, die nach Shell-Befehlen oder SQL-Fragmenten suchen, ein Zufallsgenerator. Je länger der Block, desto höher die Wahrscheinlichkeit, dass irgendwo eine Zeichenfolge auftaucht, die nach Angriff aussieht. Bei EWS stecken Anhänge als Base64 im SOAP-Body, bei Outlook im Web landen manche Aktionen als JSON im Body und teilweise sogar URL-kodiert in einem eigenen HTTP-Header. Dazu kommt Freitext: Eine Mail an die IT mit dem Satz „bitte select * from Kunden ausführen“ ist für das CRS eine SQL-Injection, und das ist aus Sicht des Regelwerks nicht einmal falsch.

Die richtige Antwort sind Ausnahmen nach Regel-ID, eingeschränkt auf Pfad und Parameter. Im CRS-Vokabular heißt das: nicht SecRuleRemoveById für den ganzen Dienst, sondern eine Laufzeit-Ausnahme, die eine Regel nur für einen bestimmten Parameter auf einem bestimmten Pfad ausblendet. Der LoadMaster erlaubt dafür eigene Regeln, die vor dem CRS laufen, gesteuert über die Option „Run First“. Das ist Handarbeit, aber genau die Handarbeit, die den Unterschied zwischen Schutz und Dekoration ausmacht.

Ausnahme-Variante	Wirkung	Wann sinnvoll
Regel global deaktivieren	Regel greift im ganzen Virtual Service nicht mehr	Nur bei Regeln, die für die Anwendung grundsätzlich unpassend sind
Regel je Pfad ausschließen	Regel fehlt nur für /owa/service.svc oder ähnliches	Standardfall für Exchange und SharePoint

Regel je Parameter ausschließen	Regel prüft einen Parameter nicht, den Rest weiter	Freitextfelder, Base64-Parameter, Anhangsinhalte
Body-Prüfung je Pfad aus	Body wird gar nicht analysiert	Große Uploads oberhalb des WAF-Maximums
WAF am SubVS aus	Kein WAF-Schutz für diesen Dienst	Binärprotokolle wie MAPI over HTTP und ActiveSync

Web-Services und Office-Clients

Bei SharePoint sprechen nicht nur Browser mit dem Server. Office-Anwendungen nutzen eigene Protokolle zum Öffnen, Sperren und Speichern von Dokumenten, dazu kommen REST über /_api und ältere SOAP-Dienste unter /_vti_bin. Diese Clients senden Header und Content-Types, die ein Browser nie verwenden würde. In der Lernphase siehst du sie an Regeln aus dem 920er-Bereich. Hier lohnt sich eine Ausnahme je Pfad mehr als eine Lockerung der Protokollregeln für den ganzen Dienst. Und wenn du ohnehin Pfade umschreibst oder intern versteckst, hilft Content Rules: Header umschreiben, Pfade umleiten, Dinge verstecken.

AUS DER PRAXIS · Health Checks, die von der eigenen WAF erschossen werden

Ein Klassiker nach dem Scharfschalten: Der Health Check eines anderen Systems, etwa eines Monitoring-Werkzeugs, ruft die VIP mit nackter IP-Adresse im Host-Header ab. Regel 920350 meldet das, und bei niedriger Schwelle steht das Monitoring plötzlich auf Rot, obwohl alles läuft. Lösung: Monitoring mit korrektem Hostnamen oder eine Ausnahme genau für dessen Quelladresse.

Geoblocking, Zugriffslisten und Rate Limiting gegen Passwort-Spraying

Die WAF prüft, was in einem Request steht. Gegen Passwort-Spraying hilft das wenig, denn eine Anmeldung mit falschem Kennwort sieht aus wie eine Anmeldung mit richtigem Kennwort. Hier kommen die anderen Werkzeuge ins Spiel: wer überhaupt anklopfen darf, von wo und wie oft.

Geoblocking: Grundrauschen senken

Der LoadMaster bietet zwei Wege. In den WAF-Einstellungen eines Virtual Service kannst du über „Countries to block“ ganze Länder sperren. Daneben gibt es eine abonnierte IP-Zugriffsliste von Progress mit bekannten bösartigen Adressen, die sich automatisch aktualisieren lässt und über die WAF-Option „IP Reputation Blocking“ wirkt. Eine eigene Ausnahmeliste hebt die Sperre für bestimmte Adressen auf. Die Daten sollten aktuell sein; ist die Liste älter als sieben Tage, weist die Oberfläche darauf hin.

Geoblocking reduziert das Grundrauschen auf OWA und ADFS spürbar. Ein großer Teil automatisierter Anmeldeversuche kommt aus Netzen, in denen ein mittelständisches Unternehmen aus dem Sauerland schlicht niemanden hat. Es ist aber kein Schutz gegen jemanden, der es ernst meint. Botnetze und gemietete Proxys mit Adressen in Deutschland kosten wenig und sind schnell beschafft.

WARNUNG · Exchange-Hybrid und Microsoft 365 nicht aussperren

In einer Exchange-Hybridumgebung greift Exchange Online auf deine lokalen EWS- und Autodiscover-Endpunkte zu, etwa für Frei/Gebucht-Abfragen und Postfachverschiebungen. Diese Zugriffe kommen aus Microsoft-Rechenzentren, und deren Adressen liegen keineswegs alle in Deutschland oder der EU. Wer Länder sperrt, ohne die veröffentlichten Exchange-Online-Bereiche auf die Ausnahmeliste zu setzen, bricht den Hybridbetrieb leise. Die aktuellen Bereiche veröffentlicht Microsoft unter Microsoft 365 URLs and IP address ranges; sie ändern sich regelmäßig, also gehört

die Pflege in einen Prozess.

Zugriffslisten je Virtual Service

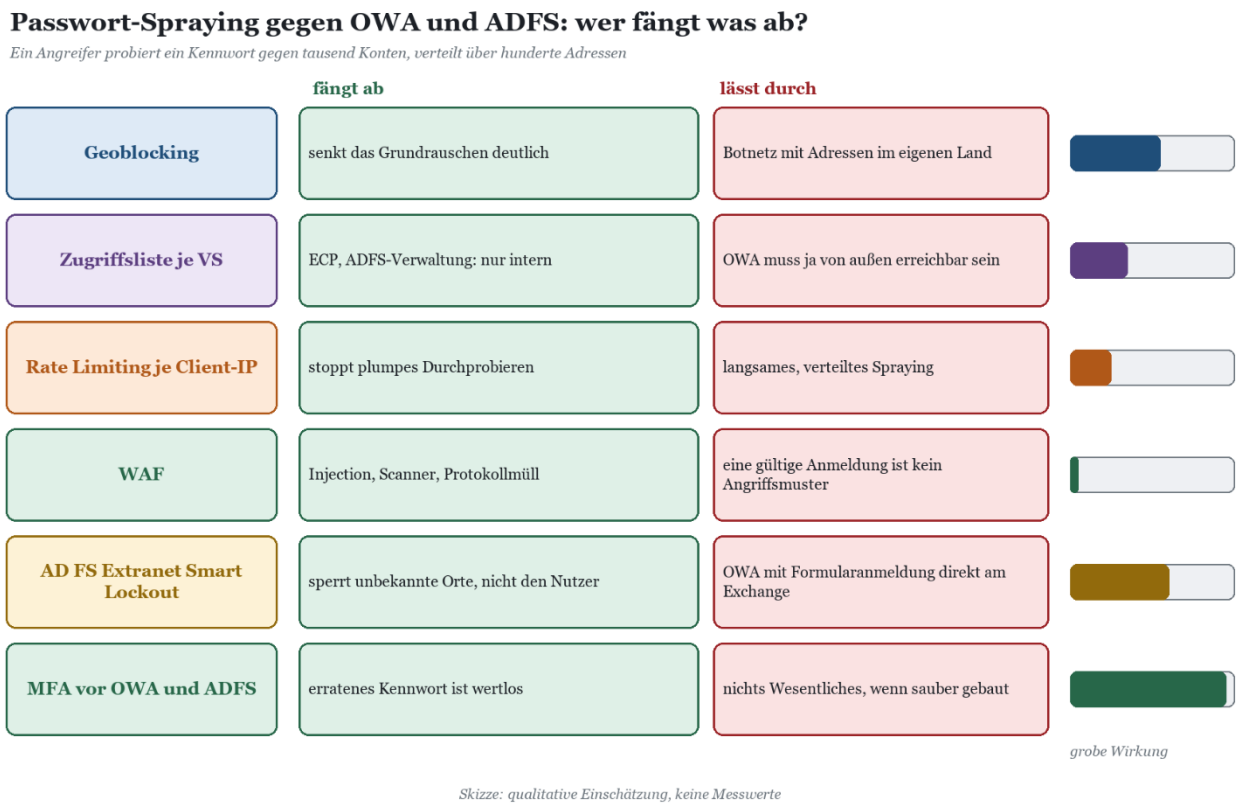
Zugriffslisten sind das stumpfste und gleichzeitig wirksamste Werkzeug: Was nicht erreichbar ist, kann nicht angegriffen werden. Der LoadMaster kennt globale Sperr- und Erlaubnislisten, die für alle Virtual Services gelten, und Zugriffslisten je Virtual Service. Zwei Eigenheiten solltest du kennen: Die Listen je Virtual Service gelten für den Haupt-Service und nicht getrennt für einzelne SubVS, und abgewiesene Verbindungen tauchen nicht in den normalen Logs auf. Wenn also jemand meldet, er komme nicht drauf, und im Log steht nichts, ist eine Zugriffsliste ein heißer Kandidat.

Weil die Liste am Haupt-Service hängt, sperrst du ECP oder die ADFS-Verwaltung nicht per SubVS-Liste, sondern über eine Content Rule, die nach Pfad und Quelladresse entscheidet, oder über einen eigenen Virtual Service mit eigenem Namen, der von außen gar nicht erst veröffentlicht wird. Letzteres ist meist die sauberere Lösung.

Rate Limiting: zählt pro Adresse, nicht pro Konto

Seit Firmware 7.2.52 bietet der LoadMaster ein überarbeitetes Client Limiting. Du kannst je Client-Adresse oder Netz drei Grenzen setzen: gleichzeitige Verbindungen, Verbindungen pro Sekunde und HTTP-Requests pro Sekunde. Dazu kommen globale Grenzen und Grenzen je Virtual Service; es greift jeweils die niedrigste, die zuerst erreicht wird. Bevor du Grenzen für einzelne Adressen setzt, muss ein globaler Standardwert existieren, und ein Client-Wert darf den globalen nicht überschreiten. Wer das HTTP-Limit reißt, bekommt eine Antwort mit Status 429; per API lässt sich stattdessen 503 einstellen.

Gegen einen einzelnen Rechner, der im Sekundentakt Kennwörter gegen OWA feuert, ist das wirksam. Gegen modernes Passwort-Spraying kaum. Der Angreifer probiert ein Kennwort gegen viele Konten, langsam und verteilt über hunderte Adressen. Jede einzelne Adresse bleibt unter jeder vernünftigen Grenze. Das Rate Limiting zählt Adressen, aber das Problem sind Konten. Und hier liegt die eigentliche Gefahr: Wird die Grenze zu niedrig gewählt, trifft sie zuerst deine eigenen Leute.



Szenario	Was der LoadMaster sieht	Folge bei strengen Grenzen
Firmen-VPN mit zentralem Internetausgang	Hunderte Nutzer, eine Adresse	Rate Limit greift beim Morgenansturm, Outlook hängt
Mobilfunk mit Carrier-NAT	Fremde Kunden teilen sich Adressen mit deinen Leuten	Sporadische 429-Fehler auf Smartphones
Reisende Mitarbeiterinnen und Mitarbeiter	Hotel-WLAN in gesperrtem Land	Kein OWA, kein ADFS, Anruf beim Helpdesk aus der Lobby
Dienstleister mit eigenem VPN	Adresse in einem Land, das du nicht erwartest	Fernwartung scheitert, Ausnahmeliste nötig
Outlook mit vielen Postfächern	Viele parallele Verbindungen von einem Client	Verbindungsgrenze reißt bei Freigaben und Gruppenpostfächern
LoadMaster hinter weiterem NAT	Nur die Adresse des vorgelagerten Systems	Geoblocking und Limits wirken auf alle gleichzeitig oder gar nicht

Die letzte Zeile ist die häufigste Ursache für völlig unerklärliches Verhalten. Geoblocking und Rate Limiting brauchen die echte Client-Adresse. Steht vor dem LoadMaster noch eine Firewall, die per Source-NAT übersetzt, sieht der LoadMaster nur sie, und jedes Limit gilt plötzlich für die gesamte Belegschaft auf einmal. Wie du die Kette sauber durchreichst, beschreibt Client-IP sichtbar machen: Transparenz, SNAT und X-Forwarded-For.

OWA und ADFS: wo Sperren wirklich hingehören

Für ADFS gibt es ein besseres Werkzeug als Rate Limiting: Extranet Smart Lockout. Seit AD FS unter Windows Server 2016 unterscheidet ADFS zwischen vertrauten und unbekannten Orten. Fehlversuche von unbekannten Adressen führen zur Extranet-Sperre, während die Anwenderin aus dem gewohnten Netz sich weiter anmelden kann. Der Schwellenwert muss dabei unter der Kontosperrschwelle des Active Directory liegen, sonst sperrt der Angreifer das Konto trotzdem

domänenweit. Eingeschaltet wird das per PowerShell, etwa mit Set-AdfsProperties -EnableExtranetLockout $true und einem passenden -ExtranetLockoutThreshold; der Modus AdfsSmartLockoutLogOnly eignet sich als eigene kleine Lernphase. Damit Smart Lockout funktioniert, muss ADFS die echte Client-Adresse kennen, was hinter Web Application Proxy und LoadMaster sauber durchgereicht werden will; Details stehen in ADFS und Web Application Proxy mit dem LoadMaster: Farm, Probe, SNI.

Outlook im Web mit Formularanmeldung direkt am Exchange hat kein Gegenstück dazu. Jeder Fehlversuch zählt gegen die AD-Kontosperre, und ein Angreifer kann damit gezielt Postfächer aussperren. Der wirksamste Schritt ist deshalb nicht eine strengere Grenze am LoadMaster, sondern eine Vorauthentifizierung mit zweitem Faktor davor. Dann ist ein erratenes Kennwort wertlos, und das Spraying läuft ins Leere. Wie das am LoadMaster aussieht, zeigen MFA vor OWA, ECP und Co.: Zwei-Faktor am LoadMaster und Edge Security Pack: Vorauthentifizierung, SSO-Domänen und der TMG-Ersatz.

TIPP · Realistische Startwerte statt Bauchgefühl

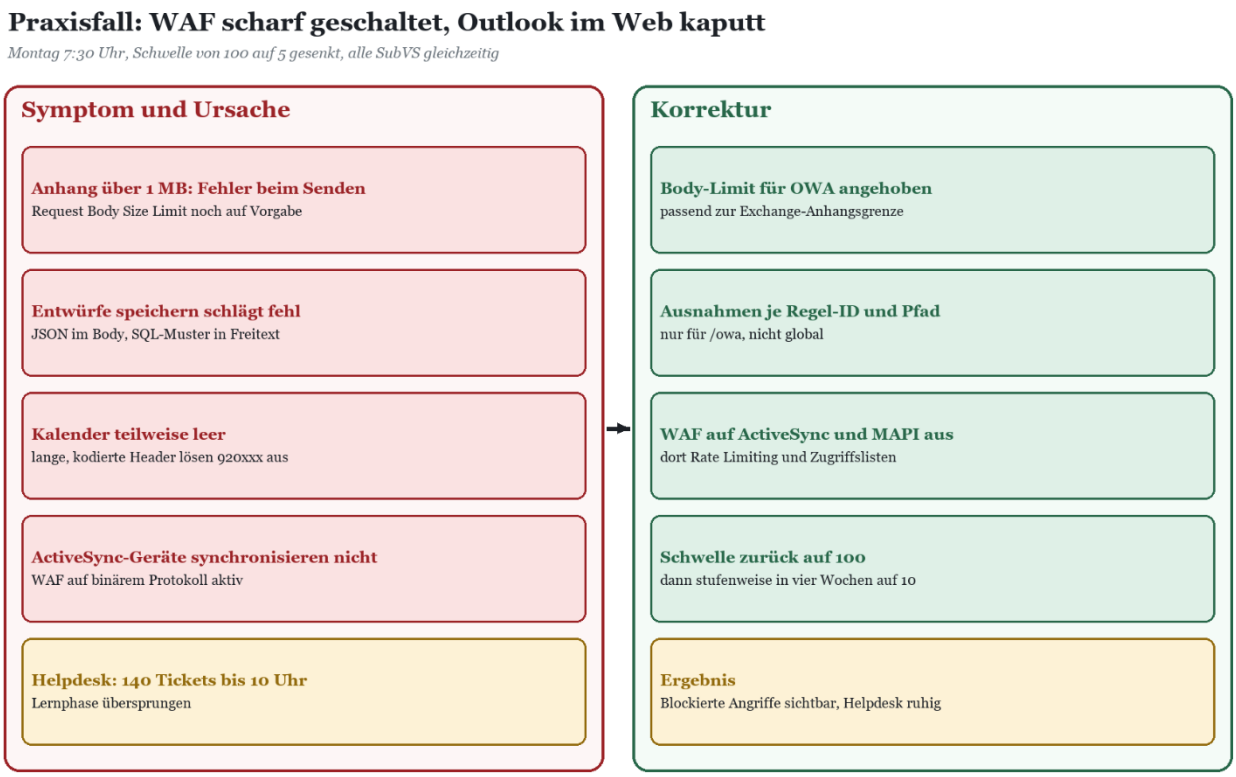
Bevor du Grenzen setzt, schau dir die Statistik des Client Limiting über eine normale Woche an. Notiere die Spitzenwerte der größten legitimen Quellen, typischerweise VPN-Ausgang und Standortanbindungen, setze diese auf eine eigene, großzügige Liste und lege die allgemeine Grenze deutlich über den typischen Wert eines einzelnen Clients. Eine Grenze, die nie greift, schadet nicht. Eine, die am Montagmorgen greift, schon.

Praxisfall: WAF scharf geschaltet, Outlook im Web kaputt

Ein mittelständisches Unternehmen mit rund 600 Postfächern, Exchange Server lokal, ein LoadMaster-Paar mit WAF-Lizenz. Die WAF lief seit Monaten mit Vorgabe-Schwelle, ein externer Prüfbericht bemängelte zu Recht, dass sie praktisch nichts blockiert. Also wurde an einem Montag um 7:30 Uhr die Schwelle für alle SubVS auf 5 gesetzt. Um 7:45 Uhr kam der erste Anruf. Um 10 Uhr lagen 140 Tickets im System, und die Geschäftsführung hatte gelernt, wie die Abkürzung WAF ausgeschrieben wird.

Was kaputt war und warum

Die Symptome waren bunt, die Ursachen überschaubar. Anhänge über 1 MB scheiterten am Request Body Size Limit, das niemand angepasst hatte, weil es bei Schwelle 100 nie aufgefallen war. Entwürfe ließen sich nicht speichern, weil Regeln aus dem 942er-Bereich im JSON-Body Freitext als SQL-Fragment lasen. Teile des Kalenders blieben leer, weil lange, kodierte Header Protokollregeln auslösten. Und die ActiveSync-Geräte der Außendienstler synchronisierten überhaupt nicht mehr, weil die WAF dort aktiv war, wo sie laut Hersteller gar nicht hingehört.



Skizze: anonymisierter Praxisfall, Namen und Zahlen verfremdet

Die Lösung Schritt für Schritt

- **Sofortmaßnahme:** Schwelle für alle SubVS zurück auf 100, Ruhe im Helpdesk binnen Minuten. Der Rückweg war dokumentiert, das hat an diesem Tag den Hals gerettet.
- **Struktur:** WAF auf ActiveSync und MAPI over HTTP abgeschaltet, dort Client Limiting und Zugriffslisten als Ausgleich. ECP extern über einen eigenen, nicht veröffentlichten Namen nur noch intern erreichbar.
- **Body-Limit:** Für das OWA-SubVS passend zur Exchange-Anhangsgrenze plus Kodierungsaufschlag angehoben, Body-Prüfung für JSON aktiv gelassen.
- **Ausnahmen:** Aus zwei Wochen Audit-Log die zwölf häufigsten Regel-IDs für /owa ausgewertet. Acht davon als Ausnahme je Pfad und Parameter, zwei als Ausnahme je Pfad, zwei blieben aktiv, weil sie echte Scanner erwischten.
- **Stufenplan:** Schwelle wöchentlich 50, 20, 10, zuerst auf ECP, dann EWS, dann OWA. Bei 10 ist das Unternehmen geblieben, weil die letzte Stufe nur noch Fehlalarme brachte.
- **Begleitung:** MFA-Vorauthentifizierung vor OWA eingeführt, AD FS Smart Lockout eingeschaltet, Geoblocking mit Ausnahmeliste für Exchange-Online-Adressen.

Kriterium	Vor dem Projekt	Nach dem Projekt
Blockierte Angriffe pro Woche	praktisch keine	mehrere hundert, überwiegend Scanner
Fehlalarme mit Ticket	keine, weil nichts blockte	vereinzelt, binnen eines Tages behoben
Anmeldeversuche auf OWA aus Ausland	Tausende pro Tag	nur noch aus Ausnahmeliste
Kontosperrungen durch Spraying	regelmäßig	seit MFA und Smart Lockout keine
Rückweg bei Störung	nicht dokumentiert	eine Minute, im Betriebshandbuch

AUS DER PRAXIS · Was wir aus dem Fall mitnehmen

Die WAF war nicht falsch konfiguriert, sie war nie eingeführt worden. Eine Schwelle von 100 ist keine Konfiguration, sondern ein Wartezimmer. Wer aus dem Wartezimmer direkt in den Operationssaal springt, sollte sich über Komplikationen nicht wundern.

FAQ: Kemp WAF, Geoblocking und Rate Limiting

Ersetzt die Kemp WAF ein Patchmanagement für Exchange?

Nein. Eine WAF kann bekannte Angriffsmuster abfangen und Zeit verschaffen, aber sie kennt nicht jede Lücke, und viele Exchange-Schwachstellen der letzten Jahre steckten in Anfragen, die für ein generisches Regelwerk unauffällig aussahen. Aktuelle Updates bleiben Pflicht.

Kann ich die WAF auf dem Haupt-Service und den SubVS gleichzeitig aktivieren?

Das solltest du nicht tun. Progress empfiehlt, die WAF auf einer Ebene zu konfigurieren, bei Exchange auf der SubVS-Ebene. Sonst wird doppelt geprüft, und die Fehlersuche wird zum Ratespiel.

Welche Schwelle ist für Outlook im Web realistisch?

Viele Umgebungen landen nach sauberer Lernphase zwischen 5 und 20. Entscheidend ist nicht die Zahl, sondern dass du sie von oben nach unten erreicht hast und jede Stufe beobachtet wurde.

Warum sehe ich abgewiesene Verbindungen nicht im Log?

Weil Zugriffslisten Verbindungen sehr früh verwerfen, bevor sie in den normalen Verbindungsprotokollen landen. Prüfe bei unerklärlichen Zugriffsproblemen deshalb zuerst die globalen Listen, die Liste des Virtual Service und die Ländersperre der WAF. Weitere Wege zur Fehlersuche stehen in Troubleshooting am LoadMaster: Logs, L7-Debug und TCP-Dump.

Hilft Geoblocking gegen Passwort-Spraying?

Es senkt die Menge automatisierter Versuche deutlich, stoppt aber keinen Angreifer, der Adressen im eigenen Land nutzt. Gegen Spraying helfen MFA und Smart Lockout, Geoblocking ist die Lärmschutzwand davor.

Wie gehe ich mit reisenden Mitarbeiterinnen und Mitarbeitern um?

Entweder verbinden sie sich zuerst mit dem Firmen-VPN, dessen Ausgang auf der Ausnahmeliste steht, oder du lässt Länder für die Dauer einer Reise frei. Die erste Variante ist planbarer, die zweite braucht einen Prozess, der nicht nur aus einem Zettel am Monitor besteht.

Bremst die Body-Prüfung den LoadMaster aus?

Sie kostet Rechenleistung und Speicher, besonders bei großen Bodies und mit Antwortprüfung. Bei Exchange mit einigen hundert Postfächern ist das auf passend dimensionierten Geräten selten ein Problem. Beobachte CPU-Last und Latenz während der Lernphase, dann weißt du es genau.

Darf ich eine Regel einfach per Klick im Log deaktivieren?

Technisch ja, fachlich nur als letzte Möglichkeit. Der Klick wirkt für den ganzen Virtual Service. Eine Ausnahme je Pfad oder Parameter hält den Schutz für den Rest aufrecht und ist die Mühe fast immer wert.

Fazit: erst zuhören, dann zuschlagen

Die Kemp WAF ist ein solides Werkzeug, solange du sie als Projekt behandelst und nicht als Häkchen. Das CRS mit Anomaly Scoring gibt dir einen Lautstärkeregler, die Paranoia Levels einen Blick in die Zukunft, das Audit-Log die Wahrheit. Exchange und SharePoint brauchen Ausnahmen,

und zwar enge: je SubVS, je Pfad, je Parameter. Binärprotokolle wie ActiveSync und MAPI over HTTP bekommen keine WAF, sondern Zugriffslisten und Grenzen.

Geoblocking und Rate Limiting sind nützlich, aber sie zählen Adressen, nicht Konten. Gegen Passwort-Spraying gewinnst du mit MFA vor OWA und ADFS und mit Extranet Smart Lockout, nicht mit immer strengeren Limits, die am Ende den Außendienst und den VPN-Ausgang treffen. Und bei allem gilt: Ohne echte Client-Adresse ist jeder Filter ein Würfelspiel. Wer das beherzigt, hat eine WAF, die Angreifer stört und die eigenen Leute in Ruhe lässt. Das ist schon mehr, als die meisten Umgebungen von sich behaupten können.

TIPP · Unterstützung bei Planung und Umsetzung

Wenn du die WAF vor Exchange, SharePoint oder ADFS einführen willst, ohne dass der Helpdesk es als Erstes merkt, oder wenn eine bestehende WAF seit Jahren im Wartezimmer sitzt, hilft dir die Beratung zum Kemp LoadMaster. Wer Audit-Logs lesen, Ausnahmen bauen und Schwellen drehen mit dem eigenen Team im Labor üben möchte, findet das in der Kemp-LoadMaster-Schulung.

WEITERLESEN · Weiter im Thema

› Content Rules: Header umschreiben, Pfade umleiten, Dinge verstecken
› MFA vor OWA, ECP und Co.: Zwei-Faktor am LoadMaster
› TLS-Offload, Re-Encryption und Zertifikate: PKI trifft LoadMaster
› Exchange Server hinter dem LoadMaster: Templates, Namespaces, Health Checks
› ADFS und Web Application Proxy mit dem LoadMaster: Farm, Probe, SNI
› Client-IP sichtbar machen: Transparenz, SNAT und X-Forwarded-For